DEPARTMENT OF APPLIED MATHEMATICS,
UNIVERSITY COLLEGE, UNIVERSITY OF LONDON.

DRAPERS' COMPANY RESEARCH MEMOIRS.

BIOMETRIC SERIES II.

MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.

XIV. ON THE GENERAL THEORY OF SKEW CORRELATION AND NON-LINEAR REGRESSION.

BY
KARL PEARSON, F.R.S.

[WITH FIVE DIAGRAMS.]

LONDON:
PUBLISHED BY DULAU AND CO., 37, SOHO SQUARE, W.
1905.

Price Five Shillings.

THIS BOOK WAS PRESENTED
TO
UNIVERSITY COLLEGE, LONDON
IN 1943
BY
THOMAS JAMES GARSTANG, M.A., Oxon.,
IN MEMORY OF THE HAPPY AND
SUCCESSFUL CAREER IN MATHEMATICS
OF HIS SON
THOMAS ERIC GARSTANG, M.Sc.,
A MEMBER OF THE COLLEGE,
1926-32.

DEPARTMENT OF APPLIED MATHEMATICS,
UNIVERSITY COLLEGE, UNIVERSITY OF LONDON.

DRAPERS' COMPANY RESEARCH MEMOIRS.

BIOMETRIC SERIES II.

MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.

XIV. ON THE GENERAL THEORY OF SKEW CORRELATION AND NON-LINEAR REGRESSION.

BY
KARL PEARSON, F.R.S.

[WITH FIVE DIAGRAMS.]

LONDON:
PUBLISHED BY DULAU AND CO., 37, SOHO SQUARE, W.
1905.

Price Five Shillings.

In March, 1903, *the Worshipful Company of Drapers announced their intention of granting* £1,000 *to the University of London to be devoted to the furtherance of research and higher work at University College. After consultation between the University and College authorities, the Drapers' Company presented* £1,000 *to the University to assist the statistical work and higher teaching of the Department of Applied Mathematics. It seemed desirable to commemorate this—probably, first occasion on which a great City Company has directly endowed higher research work in mathematical science—by the issue of a special series of memoirs in the preparation of which the Department has been largely assisted by the grant. Such is the aim of the present series of "Drapers' Company Research Memoirs."*

K. P.

Mathematical Contributions to the Theory of Evolution.—XIV. *On the General Theory of Skew Correlation and Non-linear Regression.*

By Karl Pearson, *F.R.S.*

Contents.

	Page
(1.) Introductory. General conceptions as to skew variation and correlation. General theory of skew variation within the limits of practical errors of sampling. . . .	3
(2.) Generalised idea of correlation. The correlation ratio η and its relation to the correlation coefficient r	9
(3.) Probable errors of the correlation ratio and other constants of the arrays. Probable error of r	11
(4.) On the higher types of regression. Homoscedastic and heteroscedastic systems. Homoclitic and heteroclitic systems	21
(5.) Cubical regression. General equations for regression of any order	23
(6.) Parabolic regression	28
(7.) Linear regression	30
(8.) Illustration A.—On the skew correlation between number of branches to the whorl and position of the whorl on the spray in the case of *Asperula odorata*	31
(9.) Illustration B.—On the skew correlation between age and head height in girls. . . .	34
(10.) Illustration C.—On the skew correlation between size of cell and size of body in *Daphnia magna*	38
(11.) Illustration D.—On the skew correlation between number of branches to the whorl and position of the whorl on the stem in *Equisetum arvense*.	42
(12.) Quartic regression. Necessary criteria for various types of regression	47
(13.) Illustration E.—Calculation of quartic regression in the case of *Equisetum arvense* . .	49
(14.) General conclusions. Nomenclature, clitic and scedastic curves. Difference between mere curve fitting and regression calculations. Remarks on retention of decimals .	51

(1.) *Introductory.*

In a series of memoirs presented to the Royal Society I have endeavoured to show that the Gaussian-Laplace normal distribution is very far from being a general law of frequency distribution either for errors of observation* or for the distribution of deviations from type such as occur in organic populations.† It is quite true that the

* "On Errors of Judgment, &c.," 'Phil. Trans.,' A, vol. 198, pp. 235–299.

† "On Skew Variation, &c.," 'Phil. Trans.,' A, vol. 186, pp. 343–414.

normal distribution applies within certain fields with a remarkable degree of accuracy, notably in a whole series of anthropometric, particularly craniometric, observations.* In other fields it is not even approximately correct, for example in the distribution of barometric variations,† of grades of fertility and incidence of disease.‡ For such cases I have introduced a series of skew frequency curves which serve the purpose of describing the frequency of innumerable skew distributions well within the errors of random sampling. An exact test for "goodness of fit" in the case of frequency distributions has also been now provided.§

In dealing with frequency which diverges more or less conspicuously from the normal law we require to bear in mind at least three important points :—

(i.) Any expression for frequency must be a graduation formula. It is not a disadvantage, but a fundamental requisite that it should smooth off "Scheingipfeln," so far as these are irregularities within the limits of random sampling.

Hence formulæ like those provided by THIELE‖ and WUNDT'S pupils,¶ which depend upon taking enough "moments" to reproduce the complete frequency, are *à priori* fallacious. Many interpolation formulæ would do this completely, but such interpolation formulæ are not graduation formulæ.

(ii.) The graduation formula must not depend upon the calculation of constants having such a high probable error that their value is practically worthless.

Now, the probable error of high moments and products increases rapidly with their dimensions; hence there is, beyond the labour of arithmetic, a practical limit to the number of moments or products which can be effectively used in a graduation formula.

(iii.) There must be a systematic method of approaching frequency distributions, which can be applied to all cases with reasonably practical ease.

Now the immense majority, if not the totality, of frequency distributions in homogeneous material show, when the frequency is indefinitely increased, a tendency to give a smooth curve characterised by the following properties :—

(i.) The frequency starts from zero, increases slowly or rapidly to a maximum, and then falls again to zero—probably at a quite different rate—as the character for which the frequency is measured is steadily increased. This is the almost universal unimodal distribution of the frequency of homogeneous series. Homogeneity may

* 'Biometrika,' vol. I., p. 443; vol. II., p. 344; vol. III., p. 230.

† 'Phil. Trans.,' A, vol. 190, pp. 423–469.

‡ 'Phil. Trans.,' A, vol. 192, pp. 257–330; 'The Chances of Death,' vol. I., pp. 69, *et seq.*; 'Biometrika,' vol. I., p. 134 and p. 292; and for disease, 'Phil. Trans.,' A, vol. 186, pp. 390 and 407; A, vol. 197, p. 159.

§ 'Phil. Mag.,' vol. 50, 1900, pp. 157–174, and 'Biometrika,' vol. I., pp. 154–163.

‖ 'Forelaesninger over Almindelig Iagttagelslaere,' Kjöbenhavn, 1889; 'Theory of Observations,' London, 1903.

¶ WUNDT, 'Philosophische Studien.' A whole series of papers, by G. F. LIPPS and others, seems to me to quite miss the point of (i.) and (ii.) above.

for practical purposes be taken to imply unimodality, although the converse is very far from true.

(ii.) In the next place there is generally contact of the frequency curve at the extremities of the range. These characteristics at once suggest the following form of frequency curve, if $y\delta x$ measure the frequency falling between x and $x+\delta x$:—

$$dy/dx = \frac{y\,(x+a)}{\mathrm{F}\,(x)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(i.)}.$$

For in this case we have one mode only of the frequency, *i.e.*, at $x=-a$, and dy/dx will vanish when $y=0$.

But the assumption of this form, as long as $\mathrm{F}\,(x)$ is general, is itself extremely general, and it includes cases in which dy/dx may not be zero, but take any values from 0 to ∞, when $y=0$.*

Now let us assume that $\mathrm{F}\,(x)$ can be expanded by MACLAURIN'S theorem, and equals $b_0+b_1x+b_2x^2+b_3x^3+\ldots$. Then our differential equation to the frequency will be

$$\frac{1}{y}\frac{dy}{dx} = \frac{x+a}{b_0+b_1x+b_2x^2+b_3x^3+\ldots} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(ii.)}.$$

There is now absolutely no difficulty in determining the unknown constants in terms of the moments of the system. Multiply up and also by x^n, and then integrate throughout the range of frequency, we have

$$\int x^n\,(b_0+b_1x+b_2x^2+b_3x^3+\ldots)\frac{dy}{dx}\,dx = \int y\,(x+a)\,x^n dx \quad . \quad . \quad . \quad \text{(iii.)}.$$

Or, noting that $y=0$, at the ends of the range we have, with the usual notation for a total frequency N, *i.e.*,

$$\mathrm{N}\mu'_n = \int yx^n dx \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(iv.)},$$

the result by integration by parts

$$nb_0\mu'_{n-1} + (n+1)\,b_1\mu'_n + (n+2)\,b_2\mu'_{n+1} + (n+3)\,b_3\mu'_{n+2} + \ldots = -\mu'_{n+1} - a\mu'_n \quad \text{(v.)}.$$

Hence, if we write $n=0,\ 1,\ 2,\ 3\ \ldots\ s$ successively, we have $\overline{s+1}$ equations to find $a,\ b_0,\ b_1,\ b_2\ \ldots\ b_{s-1}$ in terms of the moments. For example, if we stop at b_0 we require two moments, at b_1 three moments, at b_2 four moments, at b_3 six moments, at b_4 eight moments, and at b_{s-1}, $s>2$, $2s-2$ moments.

* For example, cases in which there is a minimum frequency or antimode at $x=-a$, and dy/dx infinite at one or two values for which $y=0$, as in the frequency distributions discussed in 'Phil. Trans.,' A, vol. 186, pp. 364–5, and 'Roy. Soc. Proc.,' vol. 62, p. 287, "Cloudiness, a Novel Case of Frequency."

There is no difficulty whatever in finding the b's; we have the system of equations; where $\mu'_0=1$

$$
\begin{array}{ll}
\mu'_0 a+0\times b_0+\ \mu'_0 b_1+2\mu'_1 b_2+3\mu'_2 b_3+4\mu'_3 b_4+\ \ldots & =-\mu'_1 \\
\mu'_1 a+\ \mu'_0 b_0+2\mu'_1 b_1+3\mu'_2 b_2+4\mu'_3 b_3+5\mu'_4 b_4+\ \ldots & =-\mu'_2 \\
\mu'_2 a+2\mu'_1 b_0+3\mu'_2 b_1+4\mu'_3 b_2+5\mu'_4 b_3+6\mu'_5 b_4+\ \ldots & =-\mu'_3 \\
\mu'_3 a+3\mu'_2 b_0+4\mu'_3 b_1+5\mu'_4 b_2+6\mu'_5 b_3+7\mu'_6 b_4+\ \ldots & =-\mu'_4 \\
\mu'_4 a+4\mu'_3 b_0+5\mu'_4 b_1+6\mu'_5 b_2+7\mu'_6 b_3+8\mu'_7 b_4+\ \ldots & =-\mu'_5 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots &
\end{array}
\quad \text{(vi.)}
$$

Hence, a, b_0, b_1, b_2, b_3, ... are at once given in terms of the determinant Δ and its minors, where:

$$
\Delta=\left|\begin{array}{ccccccc}
\mu'_0, & 0, & \mu'_0, & 2\mu'_1, & 3\mu'_2, & 4\mu'_3, & \ldots \\
\mu'_1, & \mu'_0, & 2\mu'_1, & 3\mu'_2, & 4\mu'_3, & 5\mu'_4, & \ldots \\
\mu'_2, & 2\mu'_1, & 3\mu'_2, & 4\mu'_3, & 5\mu'_4, & 6\mu'_5, & \ldots \\
\mu'_3, & 3\mu'_2, & 4\mu'_3, & 5\mu'_4, & 6\mu'_5, & 7\mu'_6, & \ldots \\
\mu'_4, & 4\mu'_3, & 5\mu'_4, & 6\mu'_5, & 7\mu'_6, & 8\mu'_7, & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{array}\right| \quad \ldots \quad \text{(vii.)}
$$

The results may be simplified slightly by taking the origin at the mean, and the moments about the mean, indicating this by dropping the dashes and putting $\mu'_1=0$.

Thus we have the following series of frequency curves, the origin being the mean:—

(i.) Keeping b_0 only

$$\frac{1}{y}\frac{dy}{dx}=-x/\mu_2 \quad \ldots\ldots\ldots \quad \text{(viii.)}$$

This is the Laplace-Gaussian normal form.

(ii.) Keeping b_0, b_1 only

$$\frac{1}{y}\frac{dy}{dx}=-\frac{x+\dfrac{\mu_3}{2\mu_2}}{\mu_2+\dfrac{\mu_3}{2\mu_2}x} \quad \ldots\ldots\ldots \quad \text{(ix.)}$$

This is the Type III. curve of my memoir on skew variation.*

(iii.) Keeping b_0, b_1, b_2 only

$$\frac{1}{y}\frac{dy}{dx}=-\frac{x+\dfrac{\mu_3(\mu_4+3\mu_2^2)}{10\mu_2\mu_4-18\mu_2^2-12\mu_3^2}}{\dfrac{\mu_2(4\mu_2\mu_4-3\mu_3^2)}{10\mu_2\mu_4-18\mu_2^3-12\mu_3^2}+\dfrac{\mu_3(\mu_4+3\mu_2^2)}{10\mu_2\mu_4-18\mu_2^3-12\mu_3^2}x+\dfrac{2\mu_2\mu_4-3\mu_3^2-6\mu_2^3}{10\mu_2\mu_4-18\mu_2^3-12\mu_3^2}x^2} \quad \text{(x.)}$$

* 'Phil. Trans.,' A, vol. 186, p. 373.

This equation gave Types I.–VI. of my two memoirs on skew variation,* and provides at once the expressions

$$d = \text{distance from mode to mean} = \frac{\sigma\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)} \quad . \quad . \quad . \quad \text{(xi.)},$$

$$\text{skewness} = \frac{\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xii.)},$$

where $\sigma = \sqrt{\mu_2}$, $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$, given in my memoir on the theory of errors of observation without proof.†

There is no *theoretical* limit, however, to this process; we can from (vi.) and (vii.) express the a and b's at once in terms of determinants, and expanding obtain forms which, like the formulæ of THIELE, will fit closer and closer to the observed distribution of frequency, the more moments we take. But there are three fundamental *practical* objections to this. These are the following :—

(*a.*) Experience shows that the form (x.) suffices for certainly the great bulk of frequency distributions, *i.e.*, it describes them effectively within the limits of random sampling.

If the distribution be even approximately normal, the series in the denominator converges very rapidly, for the coefficients of every power of x vanish for moments obeying the relationships :—

$$\mu_{2s+1} = 0, \quad \mu_{2s} = (2s-1)\mu_2\mu_{2s-2},$$

which hold for a normal series.

(*b.*) The labour of arithmetic and of analysis becomes very great, if we desire to keep higher moments. If we go to b_4 we should have to calculate the first eight moments of the observations about their centroid—a by no means easy task. Further, the classification of the resulting curves and the criteria for the right one to use in a special case, although not absolutely prohibitive, if we only go as far as b_3, are for practical purposes idle in the case of taking into account b_4.

(*c.*) The probable errors of the higher moments are so large that the values found for μ_7, μ_8, &c., are quite untrustworthy, and even that for μ_6 is doubtful,‡ unless we have frequency series far larger than usually occur in actual observations. This is a strong argument against the utility of any descriptions of frequency, such as those suggested by THIELE or LIPPS, which depend upon moments higher than the fifth or sixth.

* 'Phil. Trans.,' A, vol. 186, pp. 343–414, and 'Phil. Trans.,' A, vol. 197, pp. 443–459.

† 'Phil. Trans.,' A, vol. 198, p. 277.

‡ In 'Phil. Trans.,' A, vol. 185, pp. 71–110, I have given a method of breaking up a frequency distribution into two normal series. I obtained long ago the criterion for determining whether such a resolution is possible or not. But it involves moments higher than the fifth, and the probable error of the criterion is thus so great that for practical purposes it is worthless.

The question of the probable deviations of the higher moments can be illustrated as follows, by finding the standard deviation of the moment when we take a number of random samples from a general population. Let Σ_{μ_s} be the standard deviation of μ_s, then $100\,\Sigma_{\mu_s}/\mu_s$ is the percentage variability of μ_s due to random sampling. The table below shows the increase of these percentages in the case of the moments of normal distributions, which, quite as well as any other, will illustrate the rapid increase in probable error as we use higher and higher moments. The general values of the standard deviations of some of the moments were first given by CZUBER,* then far more completely by SHEPPARD,† and a *résumé* of all the results recently in 'Biometrika.'‡

PERCENTAGE Variability in Moments due to Random Sampling when the Series is supposed to be Normal.

Moment.	500 in series.	1000 in series.
μ_2	6·3	4·5
μ_4	14·6	10·3
μ_6	30·1	21·3
μ_8	60·6	42·9

Precisely the same rapid increase takes place when we find the variabilities of the ratios μ_4/μ_2^2, μ_6/μ_2^3, μ_8/μ_2^4, &c., which are the forms in which the moments actually occur in our coefficients. In this case we have to remember that errors in the moments are correlated, but the correlations are given in the papers cited above.§ I find in this case the following series, which is almost as suggestive as the previous table.

PERCENTAGE Variabilities in Ratio of Moments due to Random Sampling, the Series being Normal.

Ratio.	500 in series.	1000 in series.
μ_4/μ_2^2	7·3	5·2
μ_6/μ_2^3	23·3	16·5
μ_8/μ_2^4	55·1	39·0

The order of this increase of percentage variability, and therefore of probable error, is the same for skew as for normal variation, and it seems therefore, with the length

* 'Theorie der Beobachtungsfehler,' S. 130, *et seq.*

† 'Phil. Trans.,' A, vol. 192, pp. 122, *et seq.*

‡ Vol. II., pp. 273–281.

§ *Ibid.*, p. 277.

of the series in customary use, idle to use the 7^{th} or 8^{th} moments; these have variabilities varying from 30 to 60 per cent. of their values, and accordingly we might easily on a random sample reach a 7^{th} or 8^{th} moment having half, or double the value it actually has in the general population. Constants based on these high moments will be practically idle. They may enable us to describe closely an individual random sample, but no safe argument can be drawn from this individual sample as to the general population at large, at any rate so far as the argument is based on the constants depending upon these high moments.

It seems to me accordingly obvious that, bearing in mind the object of a theory of frequency (*i.e.*, the description of the distribution in the general population by aid of a *graduated* sample, agreeing with the general population within the probable errors of random sampling), we can dismiss from practical use all theories which call upon us to use moments as high as the seventh or eighth. Any use of the general form (ii.) beyond b_3, indirectly or directly, involves such higher moments. Personally I am inclined to doubt whether the continental series using higher moments are, from the standpoint of graduation, nearly as good as my form (ii.).

Hence we seem driven to the skew curves embraced in (x.) as a practical frequency series. If we have a frequency not described by (x.) we may, perhaps, use μ_5 and μ_6,* but it is difficult to see how its description can possibly be bettered by the use of still higher moments. This may seem a counsel of despair; but it is very far from being so in reality when we remember that (x.) has proved its efficiency now—I might almost say, without exception—in a wide range of economic, physical, biometric, and actuarial data.

In this memoir on skew correlation I shall accordingly confine my attention, for the most part, to constants the discovery of which does not involve the use of moments or products of higher than six dimensions, judging all above this limit to be, as a rule, disqualified for practical service by the magnitude of their probable errors.

(2.) *Generalised Idea of Correlation.*

Given any two variables or characters A and B, we say that they are correlated when, with different values x of A, we do not find the same value y of B equally likely to be associated. In other words, certain values of B are relatively more likely to occur with the value x than others. The distribution of B's associated with a given value x of A is termed an x-array of B's. If N pairs of A and B are taken, and n_x of these have the character $A = x$, these n_x form the x-array of B's. This array, like any other frequency distribution, will have its mean, which we will denote by $\bar{y}_x$, and its

* Referring to equation (ii.), I propose to call curves which stop at b_q skew curves of the q^{th} order. Thus the normal curve is a skew curve of zero order; curve of Type III. is a skew curve of the 1^{st} order; Types I., II., V., and VI. are of the 2^{nd} order. I hope shortly to publish a discussion of skew curves of the 3^{rd} order to complete the practically legitimate range of such curves.

standard deviation, which we will denote by σ_{n_x}. The mean of all the B characters shall be $\bar{y}$ and their variability given by the standard deviation σ_y. Similarly $\bar{x}$, σ_x will denote the mean and standard deviation of the A's, and n_y, $\bar{x}_y$, and σ_{n_y} the number of individuals, the mean and the standard deviation for a y-array of A's.

Now clearly a knowledge of $\bar{y}_x$ and σ_{n_x} will not fix the B's which will be found associated with a given A, but it will define the limits of probable or even possible B's. The curve obtained by plotting $\bar{y}_x$ to x is termed the regression curve of y on x. A curve in which the ratio of σ_{n_x} to the standard deviation σ_y is plotted to x may be termed a scedastic* curve. Since the standard deviation is always a positive quantity, this curve always lies on one side of the axis; it is a horizontal line in the case of normal correlation—*i.e.*, the Gauss-Laplacian distribution of deviations—and coincides with the axis, in any case where correlation passes into causation, *i.e.*, when one value of B only is associated with each A.

The mean ordinate of this curve would clearly be a sort of general measure of the degree of correlation between A and B, but it seems for many reasons better to base our measure on the mean square of the weighted standard deviations of the arrays, or

$$\sigma_{a_y}^2 = \mathrm{S}(n_x \sigma_{n_x}^2)/\mathrm{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiii.)}.$$

σ_{a_y} will thus measure the average variability in B to be found associated with any A, its vanishing will mean that the scedastic curve as defined above will coincide with the axis. Now let a new quantity η, defined by

$$\sigma_{a_y}^2 = (1-\eta^2)\,\sigma_y^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiv.)},$$

be introduced. Then clearly η must lie between ± 1, because $\sigma_{a_y}^2$ cannot be negative, being the sum of a number of positive squares. I term η the *correlation ratio*, to distinguish it from the *correlation coefficient* represented by r. When $\eta = \pm 1$ the correlation is perfect or we have causation. Further we have by a well-known property of moments, if

$$\sigma_{m_y}^2 = \mathrm{S}\{n_x (y_{n_x} - \bar{y})^2\}/\mathrm{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xv.)},$$

$$\sigma_y^2 = \sigma_{a_y}^2 + \sigma_{m_y}^2,$$

or

$$\eta = \sigma_{m_y}/\sigma_y \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xvi.)}.$$

This shows us that the correlation ratio is the ratio of the variability of the means of the x-arrays to the variability of B's in general. If $\eta = 0$, it follows that σ_{m_y} is zero, or from (xv.) that every $y_{n_x} = \bar{y}$, *i.e.*, there is no association of B's with special A's at all, or correlation is zero. Thus the correlation ratio η, as defined by either (xiv.) or (xvi.), is an excellent measure of the stringency of correlation, always lying numerically between the values 0 and 1, which mark absolute independence and

* *I.e.*, a curve which measures the "scatter" in the arrays.

complete causation respectively. Further, remembering the definition of r, the coefficient of correlation, *i.e.*,

$$\begin{aligned} N\sigma_x\sigma_y \times r &= S\{n_{xy}(x-\bar{x})(y-\bar{y})\}, \\ &= S\{n_x(x-\bar{x})(y_{n_x}-\bar{y})\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xvii.)}, \end{aligned}$$

we have, from (xv.) and (xvii.),

$$N(\eta^2-r^2)\sigma_y^2 = S\left[n_x(y_{n_x}-\bar{y})\{y_{n_x}-\bar{y}-\frac{r\sigma_y}{\sigma_x}(x-\bar{x})\}\right].$$

Now let

$$Y = \bar{y}+\frac{r\sigma_y}{\sigma_x}(x-\bar{x}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xviii.)},$$

then (xviii.), as is well known, gives the best fitting straight line to the series of points y_{n_x} loaded with their respective n_x. We can now write

$$N(\eta^2-r^2)\sigma_y^2 = S\{n_x(y_{n_x}-Y)^2\}+S\{n_x(Y-\bar{y})(y_{n_x}-Y)\}.$$

But, using (xviii.),

$$\begin{aligned} S\{n_x(Y-\bar{y})(y_{n_x}-Y)\} &= \frac{r\sigma_y}{\sigma_x}S\left[n_x(x-\bar{x})\{y_{n_x}-\bar{y}-\frac{r\sigma_y}{\sigma_x}(x-\bar{x})\}\right], \\ &= \frac{r\sigma_y}{\sigma_x}\left(Nr\sigma_x\sigma_y-\frac{r\sigma_y}{\sigma_x}N\sigma_x^2\right), \\ &= 0. \end{aligned}$$

Thus the last summation vanishes, and we have

$$N(\eta^2-r^2)\sigma_y^2 = S\{n_x(y_{n_x}-Y)^2\} \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xix.)}.$$

The right-hand side must always be positive, unless $y_{n_x}=Y$, when it is zero. Hence we conclude that η is always greater than r, or the correlation ratio greater than the correlation coefficient, except in the special case when the means of the x-arrays of y's all fall on a straight line, *i.e.*, we have linear regression, and then the two correlation constants are equal.

Thus the expression $(\eta^2-r^2)\sigma_y^2$ has an important physical meaning; it is the mean square deviation of the regression curve from the straight line which fits this curve most closely.* We have now freed our treatment of correlation from any condition as to linearity of the regression, and it remains to consider the probable errors of the various quantities dealt with.

(3.) *Probable Errors of Constants of Correlation.*

We shall first prove a number of general propositions relating to the probable errors of correlation constants. We first note that if n and n' be the frequencies in

* The properties of the correlation ratio were briefly noted in a footnote to a paper by the author in 'Roy. Soc. Proc.,' vol. 71, pp. 303–4. It has been systematically used in my laboratory for some years and determined longside r for many distributions.

any two sub-groups of a total N, for which no member of n is a member of n', then the standard deviation of n due to random sampling is given by

$$\Sigma_n^2 = n\left(1-\frac{n}{N}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (xx.),$$

and the correlation between deviations in n and n' due to random sampling is given by

$$R_{nn'}\Sigma_n\Sigma_{n'} = -\frac{nn'}{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (xxi.).$$

Problem I.—*To find the correlation in deviations due to random sampling between the number n_{x_p} in the x_p-array of y's and the number n_{y_s} in the y_s-array of x's.*

If the symbol δn denote the error or deviation in n, we have with an obvious subscript notation*

$$\delta n_{x_p} = \delta n_{x_p y_1} + \delta n_{x_p y_2} + \delta n_{x_p y_3} + \ldots + \delta n_{x_p y_q}$$

if there be q groups of y's, and again

$$\delta n_{y_s} = \delta n_{x_1 y_s} + \delta n_{x_2 y_s} + \delta n_{x_3 y_s} + \ldots + \delta n_{x_i y_s},$$

if there be i groups of x's.

Multiply the expressions for δn_{x_p} and δn_{y_s} together and we have

$$\delta n_{x_p}\delta n_{y_s} = (\delta n_{x_p y_s})^2 + S\,(\delta n_{x_p y_u}\delta n_{x_v y_s}),$$

where the summation is for every pair of values of u and v, differing from s and p.

Summing all such pairs of values for every random sample and dividing by the number of samples taken, we have the usual definition of correlation

$$\Sigma_{n_{x_p}}\Sigma_{n_{y_s}}R_{n_{x_p}n_{y_s}} = n_{x_p y_s}\left(1-\frac{n_{x_p y_s}}{N}\right) - S\left(\frac{n_{x_p y_u}n_{x_v y_s}}{N}\right);$$

or,

$$\Sigma_{n_{x_p}}\Sigma_{n_{y_s}}R_{n_{x_p}n_{y_s}} = n_{x_p y_s} - \frac{n_{x_p}n_{y_s}}{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (xxii.).$$

This gives $R_{n_{x_p}n_{y_s}}$, the required correlation, since $\Sigma_{n_{x_p}}$ and $\Sigma_{n_{y_s}}$ are known from (xx.).

Problem II.—*To find the correlation between deviations in the total n_{x_p} of any array and in any sub-group $n_{x_p y_s}$ of this array.*

We have at once

$$\delta n_{x_p}\delta n_{x_p y_s} = (\delta n_{x_p y_s})^2 + S\,(\delta n_{x_p y_u}\delta n_{x_p y_s})$$

where u is to be taken every value other than s in the summation term. Summing for all random samples and dividing by their number, we have, after using results like (xx.) and (xxi.),

$$R_{n_{x_p}n_{x_p y_s}} \times \Sigma_{n_{x_p}}\Sigma_{n_{x_p y_s}} = n_{x_p y_s}\left(1-\frac{n_{x_p}}{N}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (xxiii.),$$

which gives $R_{n_{x_p}n_{x_p y_s}}$.

* n_{xy} = frequency of groups with characters x and y.

Proposition III.—*There is no correlation between deviations in the mean of an x-array* y_{x_p} *and the total number in that array.*

$$n_{x_p} \times y_{x_p} = \mathrm{S}\,(n_{x_p y_u} y_u),$$

$$n_{x_p} \delta y_{x_p} = \mathrm{S}\,(\delta n_{x_p y_u} y_u) - y_{x_p} \delta n_{x_p},$$

$$n_{x_p} \delta y_{x_p} \delta n_{x_p} = -y_{x_p} (\delta n_{x_p})^2 + \mathrm{S}\,(\delta n_{x_p} \delta n_{x_p y_u} y_u).$$

Hence as before, using (xxiii.), &c.,

$$\begin{aligned} n_{x_p} \Sigma_{y_{x_p}} \Sigma_{n_{x_p}} \mathrm{R}_{y_{x_p} n_{x_p}} &= -y_{x_p} n_{x_p} \left(1 - \frac{n_{x_p}}{\mathrm{N}}\right) + \mathrm{S}\left\{ n_{x_p y_u} \left(1 - \frac{n_{x_p}}{\mathrm{N}}\right) y_u \right\} \\ &= -y_{x_p} n_{x_p} \left(1 - \frac{n_{x_p}}{\mathrm{N}}\right) + \left(1 - \frac{n_{x_p}}{\mathrm{N}}\right) n_{x_p} y_{x_p} \\ &= 0, \end{aligned}$$

which proves that $\mathrm{R}_{y_{x_p} n_{x_p}}$ is zero.

Proposition IV.—*There is no correlation between deviations in the mean of an x-array and in the total number in any other array.*

Proof as before.

Proposition V.—*There is no correlation between deviations in the mean of one x-array and in the mean of a second x-array.*

We have

$$n_{x_p} \delta y_{x_p} = \mathrm{S}\,(\delta n_{x_p y_u} y_u) - y_{x_p} \delta n_{x_p},$$

$$n_{x_{p'}} \delta y_{x_{p'}} = \mathrm{S}\,(\delta n_{x_{p'} y_u} y_u) - y_{x_{p'}} \delta n_{x_{p'}}.$$

Multiply these two expressions together, sum for all random samples, and divide by the number of such samples. We find

$$\begin{aligned} n_{x_p} n_{x_{p'}} \Sigma_{y_{x_p}} \Sigma_{y_{x_{p'}}} \mathrm{R}_{y_{x_p} y_{x_{p'}}} &= -y_{x_p} y_{x_{p'}} \frac{n_{x_p} n_{x_{p'}}}{\mathrm{N}} \\ &\quad + y_{x_p} \mathrm{S}\,(n_{x_p} n_{x_{p'} y_u} y_u)/\mathrm{N} \\ &\quad + y_{x_{p'}} \mathrm{S}'\,(n_{x_{p'}} n_{x_p y_u} y_u)/\mathrm{N} \\ &\quad - \mathrm{S}\,(n_{x_p y_u} n_{x_{p'} y_u} y_u^2)/\mathrm{N} \\ &\quad - \mathrm{S}'\,(n_{x_p y_u} n_{x_{p'} y_{u'}} y_u y_{u'})/\mathrm{N} \\ &= -y_{x_p} y_{x_{p'}} \frac{n_{x_p} n_{x_{p'}}}{\mathrm{N}} + y_{x_p} \frac{n_{x_p} n_{x_{p'}}}{\mathrm{N}} y_{x_{p'}} \\ &\quad + y_{x_{p'}} \frac{n_{x_{p'}} n_{x_p}}{\mathrm{N}} y_{x_p} - \frac{\mathrm{S}(n_{x_p y_u} y_u) \times \mathrm{S}(n_{x_{p'} y_{u'}} y_{u'})}{\mathrm{N}}. \end{aligned}$$

The last term is $\frac{n_{x_p} y_{x_p} \times n_{x_{p'}} y_{x_{p'}}}{\mathrm{N}}$, and thus the right-hand side is identically zero. It thus appears that there is no correlation between errors made in finding the means of two arrays. This result is not at once obvious, although a very little consideration shows it must be true.

Proposition VI.—*To prove that the standard deviation of the mean y_{x_p} of any x-array due to random sampling equals* $\frac{\sigma_{n_{x_p}}}{\sqrt{n_{x_p}}}$.

We have

$$n_{x_p}\delta y_{x_p} = \mathrm{S}'(\delta n_{x_p y_u} y_u) - y_{x_p}\delta n_{x_p}.$$

Square, sum for all random samples, and divide by the number of such samples. We have

$$\begin{aligned} n_{x_p}^2 \Sigma_{y_{x_p}}^2 &= y_{x_p}^2 n_{x_p}\left(1-\frac{n_{x_p}}{\mathrm{N}}\right) - 2y_{x_p}\mathrm{S}\left\{n_{x_p y_u}\left(1-\frac{n_{x_p}}{\mathrm{N}}\right)y_u\right\} \\ &\quad + \mathrm{S}\left\{n_{x_p y_u}\left(1-\frac{n_{x_p y_u}}{\mathrm{N}}\right)y_u^2\right\} \\ &\quad - 2\mathrm{S}\left\{\frac{n_{x_p y_u} n_{x_p y_{u'}}}{\mathrm{N}} y_u y_{u'}\right\} \\ &= y_{x_p}^2 n_{x_p}\left(1-\frac{n_{x_p}}{\mathrm{N}}\right) - 2y_{x_p}^2 n_{x_p}\left(1-\frac{n_{x_p}}{\mathrm{N}}\right) \\ &\quad + \mathrm{S}(n_{x_p y_u} y_u^2) - \frac{\mathrm{S}(n_{x_p y_u} y_u)\,\mathrm{S}(n_{x_p y_{u'}} y_{u'})}{\mathrm{N}} \\ &= \mathrm{S}(n_{x_p y_u} y_u^2) - n_{x_p} y_{x_p}^2 \\ &= n_{x_p}\sigma_{n_{x_p}}^2. \end{aligned}$$

Hence

$$\Sigma_{y_{x_p}} = \sigma_{n_{x_p}}/\sqrt{n_{x_p}} \quad \ldots\ldots\ldots \quad \text{(xxiv.)}.$$

Thus the probable error of the mean of an array has exactly the same form as the probable error of the mean of a random sample of a *definite* number of individuals. The array may have a variable number of individuals, but we have seen in Proposition III. that there is no correlation between errors in its mean and errors in the total number of individuals contained in it.

Problem VII.—*To find the probable error of the standard deviation of any array.*

By a precisely similar investigation to that of the previous proposition we find

$$\Sigma_{\sigma_{n_{x_p}}} = \sqrt{\frac{m_4 - m_2^2}{4 n_{x_p} m_2}} \quad \ldots\ldots\ldots \quad \text{(xxv.)},$$

where

$$m_s = \frac{1}{n_{x_p}}\mathrm{S}\{(y_u - y_{x_p})^s n_{x_p y_u}\}.$$

This is identical with the probable error we should have if the array were a random sample of constant size.

In many cases it will be sufficiently approximate to put $m_4 = 3m_2^2$ and we then have

$$\cdot 67449\,\Sigma_{\sigma_{n_{x_p}}} = \cdot 67449\,\frac{\sigma_{n_{x_p}}}{\sqrt{2n_{x_p}}} \quad \ldots\ldots\ldots \quad \text{(xxvi.)},$$

the well-known form for the probable error of the standard deviation of a normal distribution of a definite number of individuals.

Problem VIII.—*To find the standard deviation of the standard-deviation* σ_M *of the means of the arrays due to random sampling.*

Since

$$N\sigma_M^2 = S\{n_{x_p}(y_{x_p}-\bar{y})^2\}$$

$$2N\sigma_M\,\delta\sigma_M = S\{\delta n_{x_p}(y_{x_p}-\bar{y})^2\} + 2S\{\delta y_{x_p} n_{x_p}(y_{x_p}-\bar{y})\} - 2\delta\bar{y}S\{n_{x_p}(y_{x_p}-\bar{y})\},$$

the last term of which vanishes, since

$$N\bar{y} = S(n_{x_p} y_{x_p}).$$

Square the above relation, sum for all random samples, and divide by the number of such samples.

We find

$$\begin{aligned}
4N^2\sigma_M^2\Sigma_{\sigma_M}^2 = {} & S\left\{n_{x_p}\left(1-\frac{n_{x_p}}{N}\right)(y_{x_p}-\bar{y})^4\right\} \\
& -2S\left\{\frac{n_{x_p}n_{x_{p'}}}{N}(y_{x_p}-\bar{y})^2(y_{x_{p'}}-\bar{y})^2\right\} \\
& +4S\{\Sigma_{n_{x_p}}\Sigma_{y_{x_p}}R_{n_{x_p}y_{x_p}}(y_{x_p}-\bar{y})^3\} \\
& +4S\{\Sigma_{n_{x_{p'}}}\Sigma_{y_{x_p}}R_{n_{x_{p'}}y_{x_p}}(y_{x_{p'}}-\bar{y})^2(y_{x_p}-\bar{y})\} \\
& +4S\{\Sigma_{y_{x_p}}\Sigma_{y_{x_p}}R_{y_{x_p}y_{x_p}}(y_{x_p}-\bar{y})(y_{x_{p'}}-\bar{y})\} \\
& +4S\{\Sigma y_{x_p}^2 n_{x_p}^2(y_{x_p}-\bar{y})^2\}.
\end{aligned}$$

But $R_{n_{x_p}y_{x_p}}$, $R_{n_{x_{p'}}y_{x_p}}$, and $R_{y_{x_p}y_{x_{p'}}}$ vanish by Propositions III., IV., and V. Further, by VI., $\Sigma_{y_{x_p}}^2 = \sigma_{n_{x_p}}^2/n_{x_p}$. Hence we have

$$\begin{aligned}
4N^2\sigma_M^2\Sigma_{\sigma_M}^2 = {} & S\left\{n_{x_p}\left(1-\frac{n_{x_p}}{N}\right)(y_{x_p}-\bar{y})^4\right\} \\
& -2S\left\{\frac{n_{x_p}n_{x_{p'}}}{N}(y_{x_p}-\bar{y})^2(y_{x_{p'}}-\bar{y})^2\right\} \\
& +4S\{n_{x_p}\sigma_{n_{x_p}}^2(y_{x_p}-\bar{y})^2\} \\
= {} & S\{n_{x_p}(y_{x_p}-\bar{y})^4\} - \frac{[S\{n_{x_p}(y_{x_p}-\bar{y})\}]^2}{N} \\
& +4S\{n_{x_p}\sigma_{n_{x_p}}^2(y_{x_p}-\bar{y})^2\}.
\end{aligned}$$

Now let

$$N\lambda_q = S\{n_{x_p}(y_{x_p}-\bar{y})^q\}$$

be the n^{th} moment of the means of the arrays about their mean. Then clearly $\lambda_2 = \sigma_M^2$. Further, since $S(n_{x_p}\sigma_{n_{x_p}}^2) = N\sigma_y^2(1-\eta^2)$, we can write

$$S\{n_{x_p}\sigma_{n_{x_p}}^2(y_{x_p}-\bar{y})^2\} = N\sigma_y^2(1-\eta^2)\,\sigma_M^2 \times \chi_1,$$

where χ_1 is a purely numerical constant, which is equal to unity for those cases in which there is no correlation between the standard deviation of an array and the square of its mean's deviation from the mean. Thus finally we find

$$\Sigma_{\sigma_M}{}^2=\frac{\lambda_4-\lambda_2{}^2}{4N\lambda_2}+\frac{\chi_1\sigma_y{}^2(1-\eta^2)}{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxvii.)}.$$

This enables us at once to find the probable error of the standard deviation of the means of the arrays.

Proposition IX.—*To find the correlation between the deviations due to random sampling in the values of* σ_y *and* σ_M.

We have

$$N\sigma_y{}^2=S\{n_y(y-\bar{y})^2\},$$

$$2N\sigma_y\delta\sigma_y=S\{\delta n_{y_s}(y_s-\bar{y})^2\}-2\,\delta\bar{y}\,S\{n_{y_s}(y_s-\bar{y})\}\ ;$$

the last term vanishes because $S(n_{y_s}y_s)=N\bar{y}$.

Thus

$$2N\sigma_y\delta\sigma_y=S\{\delta n_{y_s}(y_s-\bar{y})^2\}.$$

But from the previous proposition

$$2N\sigma_M\delta\sigma_M=S\{\delta n_{x_p}(y_{x_p}-\bar{y})^2\}+2S\{\delta y_{x_p}n_{x_p}(y_{x_p}-\bar{y})\}.$$

Multiply these two expressions together, sum for all random samples and divide by the number of such samples; we find

$$4N^2\sigma_y\sigma_M\Sigma_{\sigma_y}\Sigma_{\sigma_M}R_{\sigma_y\sigma_M}=S\{\Sigma_{n_{y_s}}\Sigma_{n_{x_p}}(y_s-\bar{y})^2(y_{x_p}-\bar{y})^2R_{n_{y_s}n_{x_p}})\}$$
$$+2S\{n_{x_p}\Sigma_{n_{y_s}}\Sigma_{y_{x_p}}R_{n_{y_s}y_{x_p}}(y_s-\bar{y})^2(y_{x_p}-\bar{y})\}.$$

To evaluate this, we require to find the two correlations expressed by $R_{n_{y_s}n_{x_p}}$ and $R_{n_{y_s}y_{x_p}}$. We will consider the two summation terms separately.

First Term.
$$\delta n_{x_p}=\delta n_{x_py_1}+\delta n_{x_py_2}+\ldots+\delta n_{x_py_s}+\ldots$$
$$\delta n_{y_s}=\delta n_{y_sx_1}+\delta n_{y_sx_2}+\ldots+\delta n_{y_sx_p}+\ldots$$
$$\delta n_{x_p}\delta n_{y_s}=(\delta n_{x_py_s})^2+S(\delta n_{x_py_{s'}}\delta n_{x_{p'}y_s}),$$

where in the summation p' and s' are not equal to p and s.

Proceeding in the usual manner we find

$$\Sigma_{n_{x_p}}\Sigma_{n_{y_s}}R_{n_{x_p}n_{y_s}}=n_{x_py_s}\left(1-\frac{n_{x_py_s}}{N}\right)-S\left\{\frac{n_{x_py_{s'}}n_{x_{p'}y_s}}{N}\right\}$$
$$=n_{x_py_s}-\frac{S(n_{x_py_{s'}})\times S(n_{x_{p'}y_s})}{N},$$

where in the first sum s' is to take all possible values, and in the second p' is to take all possible values. Thus we have

$$\Sigma_{n_{x_p}}\Sigma_{n_{y_s}}R_{n_{x_p}n_{y_s}}=n_{x_py_s}-\frac{n_{x_p}n_{y_s}}{N} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxviii.)}.$$

Substituting we find

$$\textit{First Term} = S_1\{n_{x_py_s}(y_s-\bar{y})^2(y_{x_p}-\bar{y})^2\} - S_2\left\{\frac{n_{x_p}n_{y_s}}{N}(y_s-\bar{y})^2(y_{x_p}-\bar{y})^2\right\}.$$

Here both the summations are really double summations; fixing our attention on any x_p, *i.e.*, on any array of y's for a given value of x, we have first to sum for all y's in this array, and then we have to sum for all arrays. This is the meaning of S_1. In S_2 we are to associate every array of x's with every array of y's; hence this term will break up at once into two factors, *i.e.*,

$$\begin{aligned}&\frac{1}{N}S\{n_{x_p}(y_{x_p}-\bar{y})^2\}\times S\{n_{y_s}(y_s-\bar{y})^2\}\\ &\quad=\sigma_y^2\times S\{n_{x_p}(y_{x_p}-\bar{y})^2\}\\ &\quad=N\sigma_y^2\times\sigma_M^2.\end{aligned}$$

Keeping x_p constant first in S_1, we see that

$$S\{n_{x_py_s}(y_s-\bar{y})^2\}$$

is the 2nd moment of the y's in the x_p array about the mean of the system

$$=n_{x_p}\{\sigma_{n_{x_p}}^2+(y_{x_p}-\bar{y})^2\}.$$

Combining we have

$$\begin{aligned}\textit{First Term} &= S\{n_{x_p}(y_{x_p}-\bar{y})^4\}+S\{n_{x_p}\sigma_{n_{x_p}}^2(y_{x_p}-\bar{y})^2\}-N\sigma_y^2\sigma_M^2\\ &= N\{\lambda_4+\sigma_y^2\sigma_M^2(1-\eta^2)\chi_1-\sigma_y^2\sigma_M^2\} \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxix.)}.\end{aligned}$$

We now turn to the second term which involves the discovery of $R_{n_{y_s}y_{x_p}}$.

$$\delta n_{y_s}\delta y_{x_p}=(\delta n_{y_sx_1}+\delta n_{y_sx_2}+\ldots+\delta n_{y_sx_p}+\ldots)\,\delta y_{x_p}$$

$$n_{x_p}\delta y_{x_p}=-y_{x_p}\delta n_{x_p}+S(\delta n_{x_py_u}y_u).$$

Hence

$$\begin{aligned}n_{x_p}\delta n_{y_s}\delta y_{x_p}=&-y_{x_p}(\delta n_{y_sx_1}+\delta n_{y_sx_2}+\ldots+\delta n_{y_sx_p}+\ldots)\,\delta n_{x_p}\\ &+(\delta n_{y_sx_1}+\delta n_{y_sx_2}+\ldots+\delta n_{y_sx_p}+\ldots)\,S(\delta n_{x_py_u}y_u).\end{aligned}$$

Sum for all random samples and divide by the number of such samples; we have

$$\begin{aligned}n_{x_p}\Sigma_{n_{y_s}}\Sigma_{y_{x_p}}R_{n_{y_s}y_{x_p}}=&-y_{x_p}\left(n_{x_p}y_s-\frac{n_{x_p}n_{y_s}}{N}\right)\\ &+n_{x_py_s}y_s-\frac{S(y_un_{x_py_u}n_{x_p})}{N}\\ =&\,n_{x_py_s}(y_s-y_{x_p}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxx.)}.\end{aligned}$$

Substituting we have

$$Second\ Term = 2S\{n_{x_p y_s}(y_s - y_{x_p})(y_s - \bar{y})^2 (y_{x_p} - \bar{y})\}.$$

Here again the summation is of a double character.

Let us first take x_p as constant and sum for every value of y_s. We may write $y_s - \bar{y} = (y_s - y_{x_p} + y_{x_p} - \bar{y})$, and our first summation will be

$$2(y_{x_p} - \bar{y}) \times S[n_{x_p y_s}\{(y_s - y_{x_p})^3 + 2(y_s - y_{x_p})^2 (y_{x_p} - \bar{y}) + (y_s - y_{x_p})(y_{x_p} - \bar{y})^2\}]$$
$$= 2(y_{x_p} - \bar{y})\, n_{x_p} m_3 + 4(y_{x_p} - \bar{y})^2\, n_{x_p} m_2 + 2(y_{x_p} - \bar{y})^3\, S\{n_{x_p y_s}(y_s - y_{x_p})\},$$

if

$$n_{x_p} m_q = S\{n_{x_p y_s}(y_s - y_{x_p})^q.$$

The last term vanishes for $S(n_{x_p y_s} y_s) = n_{x_p} y_{x_p}$ by the definition of the mean.

Hence

$$Second\ Term = 2S\{n_{x_p} m_3 (y_{x_p} - \bar{y})\} + 4S\{n_{x_p} \sigma_{n_{x_p}}{}^2 (y_{x_p} - \bar{y})^2\}.$$

Here m_3 is the third moments of the x_p array of y's, which will probably be very small if the arrays are nearly symmetrical and the first term clearly depends on the existence of a correlation between the skewness of the arrays and the magnitude of their means.

We may write the first term then :

$$= 2N\sigma_{a_y}{}^3 \sigma_M \times \chi_2$$
$$= 2N\sigma_y{}^3 (1-\eta^2)^{3/2} \sigma_M \times \chi_2,$$

where χ_2 is a purely numerical quantity, which for most cases will probably be very small or even zero.

Thus we find :

$$Second\ Term = 2N\sigma_y{}^3 (1-\eta^2)^{3/2} \sigma_M \chi_2 + 4N\sigma_y{}^2 \sigma_M{}^2 (1-\eta^2)\chi_1 \quad . \quad . \quad (xxxi.).$$

We can now return to p. 16 and write down the full correlation between deviations in the values of σ_y and σ_M due to random sampling. Remembering that $\sigma_M = \eta\sigma_y$,* we find :

$$\Sigma_{\sigma_y}\Sigma_{\sigma_M} R_{\sigma_y \sigma_M} = \frac{1}{4N\eta}\left[\frac{\lambda_4}{\sigma_y{}^2} + \eta^2 \sigma_y{}^2 \{(1-\eta^2)\chi_1 - 1\}\right]$$
$$+ \frac{1}{2N}\sigma_y{}^2 (1-\eta^2)^{3/2}\chi_2 + \frac{\sigma_y{}^2}{N}\eta(1-\eta^2)\chi_1$$
$$= \frac{\sigma_y{}^2}{N}\left\{\frac{\lambda_4}{4\eta\sigma_y{}^4} + \tfrac{5}{4}\eta(1-\eta^2)\chi_1 - \tfrac{1}{4}\eta + \tfrac{1}{2}(1-\eta^2)^{3/2}\chi_2\right\} \quad . \quad . \quad . \quad . \quad (xxxii.).$$

* It should be remembered that this definition of η gives it invariably the *positive* sign.

Proposition X.—*To find the standard deviation of the values of the correlation ratio* η *due to random sampling,* i.e., *to find the probable error of the correlation ratio* η.

We have

$$\eta = \sigma_M/\sigma_y.$$

Hence

$$\frac{\delta\eta}{\eta} = \frac{\delta\sigma_M}{\sigma_M} - \frac{\delta\sigma_y}{\sigma_y}.$$

Squaring, summing for all random samples and dividing by the number of such samples, we have :

$$\frac{\Sigma_\eta^2}{\eta^2} = \frac{\Sigma_{\sigma_M}^2}{\sigma_M^2} + \frac{\Sigma_{\sigma_y}^2}{\sigma_y^2} - \frac{2\Sigma_{\sigma_y}\Sigma_{\sigma_M}R_{\sigma_M\sigma_y}}{\sigma_M\sigma_y}.$$

$\Sigma_{\sigma_M}^2$ is given (xxvii.), $\Sigma_{\sigma_y}\Sigma_{\sigma_M}R_{\sigma_M\sigma_y}$ by (xxxii.) and $\Sigma_{\sigma_y}^2 = \frac{1}{4N}\frac{\mu_4-\mu_2^2}{\mu_2}$ by a well-known formula.*

Substituting, we have the complete value of Σ_η given by :

$$\frac{\Sigma_\eta^2}{\eta^2} = \frac{\lambda_4-\lambda_2^2}{4N\lambda_2^2} + \frac{\chi_1(1-\eta^2)}{N\eta^2} + \frac{1}{4N}\frac{\mu_4-\mu_2^2}{\mu_2^2} - \frac{1}{2N}\frac{\lambda_4}{\lambda_2^2}\eta^2 - \frac{5}{2N}(1-\eta^2)\chi_1 + \frac{1}{2N} - \frac{(1-\eta^2)^{3/2}\chi_2}{N\eta};$$

or, after re-arranging,

$$\Sigma_\eta^2 = \frac{1}{N}\left\{(1-\eta^2)^2 + \frac{\mu_4-3\mu_2^2}{4\mu_2^2}\eta^2 + \frac{\lambda_4-3\lambda_2^2}{4\lambda_2^2}\eta^2(1-2\eta^2) + (\chi_1-1)(1-\eta^2)(1-\tfrac{5}{2}\eta^2) - \chi_2\eta(1-\eta^2)^{3/2}\right\} \quad . \quad . \quad . \quad \text{(xxxiii.)}.$$

For normal correlation, $\mu_4 = 3\mu_2^2$. Further

$$y_{x_p} - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x_p - \bar{x}),$$

and

$$N\lambda_4 = S\{n_{x_p}(y_{x_p}-\bar{y})^4\} = \frac{r^4\sigma_y^4}{\sigma_x^4} S\{n_{x_p}(x_p-\bar{x})^4\}$$

$$= \frac{r^4\sigma_y^4}{\sigma_x^4} \times N3\sigma_x^4 = 3N\lambda_2^2.$$

Hence the second and third terms vanish. Further $\chi_1 = 1$ and $\chi_2 = 0$, while $\eta = r$.

Hence we have

$$\Sigma_\eta^2 = \Sigma_r^2 = \frac{(1-r^2)^2}{N},$$

which agrees with the special result.

* 'Biometrika,' vol. II., p. 276.

In any other case, χ_2, χ_1-1, $(\mu_4-3\mu_2^2)/\mu_2^2$, $(\lambda_4-3\lambda_2^2)/\lambda_2^2$ will probably be small and thus

$$\Sigma_\eta^2=\frac{1}{N}(1-\eta^2)^2.$$

Probable error of

$$\eta=\cdot67449\,(1-\eta^2)/\sqrt{N}, \text{ nearly} \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxxiv.)}.$$

This simple form suffices for many practical cases.

If greater exactitude is wanted, there is, however, no great labour in using (xxxiii.). We find the means and standard deviations of each array.

Then $N\lambda_2$ and $N\lambda_4$ are the 2nd and 4th moments of the means of these arrays about their mean.

$N\mu_2$ and $N\mu_4$ are the 2nd and 4th moments about the mean of the y-characters, and will always be known for *skew* variation.

χ_1 is defined by

$$\chi_1=\frac{S\{n_{x_p}\sigma_{n_{x_p}}^2(y_{x_p}-\bar{y})^2\}}{N\sigma_y^2(1-\eta^2)\,\sigma_M^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxxv.)},$$

and can be easily found when the means and standard deviations of each array have been found.

The most troublesome expression is χ_2 defined by

$$\chi_2=\frac{S\{n_{x_p}m_3(y_{x_p}-\bar{y})\}}{N\sigma_y^3(1-\eta^2)^{\frac{3}{2}}\sigma_M} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxxvi.)}.$$

But as we do not take usually more than 10 to 20 arrays, the discovery of their 3rd moments is not an extremely difficult task. As a rule, however, χ_2 is very small and may be fairly neglected, even when we must find χ_1-1. All these points will be dealt with in the numerical illustrations given later in this paper. At present we note that the probable error of η has been determined, and that its value for the general case is not really more complex than the value of the probable error of r in the general case, which requires the determination of product moments of the 4th order.*

* Let $Np_{qs}=S\{n_{xy}(x-\bar{x})^q(y-\bar{y})^s\}$, then the probable error of r is given by

$$\Sigma_r^2=\frac{r^2}{N}\left\{\frac{p_{22}-3p_{11}^2}{p_{11}^2}+\frac{p_{22}-3p_{20}p_{02}}{2p_{20}p_{02}}+\frac{p_{40}-3p_{20}^2}{4p_{20}^2}+\frac{p_{04}-3p_{02}^2}{4p_{02}^2}-\frac{p_{31}-3p_{11}p_{20}}{p_{11}p_{20}}-\frac{p_{13}-3p_{11}p_{02}}{p_{11}p_{02}}\right\}. \quad \text{(xxxvii.)}.$$

This agrees with the value given by SHEPPARD ('Phil. Trans.,' A, vol. 192, p. 128), except that the r^2 factor has been dropped by a printer's error in his paper. For the special case of a normal distribution, we have easily from the equation to the normal surface

$$p_{40}=3p_{20}^2,\quad p_{04}=3p_{02}^2,\quad p_{31}=3p_{11}p_{20},\quad p_{13}=3p_{11}p_{02},\quad (p_{22}-3p_{11}^2)/p_{11}^2=(1-r^2)/r^2$$

and

$$\frac{p_{22}-3p_{20}p_{02}}{2p_{20}p_{02}}=r^2-1, \qquad \text{whence} \quad \Sigma_r=(1-r^2)/\sqrt{N},$$

the well-known form ('Phil. Trans.,' A, vol. 191, p. 245).

(4.) *On the Higher Types of Regression.*

We have already seen how the introduction of the correlation ratio η enables us to drop the limitations associated with the Gauss-Laplacian form of frequency, and the Bravais correlation formulæ. The fundamental step towards this advance was undoubtedly taken by G. U. YULE in his paper in the 'Roy. Soc. Proc.,' vol. 60, pp. 477 *et seq.*, wherein he shows that if the regression be linear, the Bravais type of formula applied to multiple correlation is still true, although we make no assumption as to the form of the frequency surface. It would undoubtedly be a gain to have skew frequency surfaces which would describe skew correlation for the great mass of cases as effectivly as the series of skew frequency curves describe skew variation, but although a considerable amount of progress has been made in the consideration of these surfaces, their full theory has not yet been worked out owing to difficulties of analysis, and their complete discussion must still be postponed. YULE'S method of approaching the problem from the form of the regression curves is, however, available and capable of very great extension. Its chief advantage is that it makes little or no assumption as to the distribution of frequency; its chief defect lies even in this advantage of generality: it does not enable us to predict the probability of an individual with a given combination of characters. This follows at once from the fact that we make no assumption as to the form of the distribution within an array. Without some theory as to variation within the array, we are reduced to the laborious process of calculating the standard deviation, skewness, and other general characters of each array, a lengthy and troublesome process compared with a theory which would, like the Bravais theory, give these at once in terms of a few constants determined from the data as a whole.

In the great bulk of biometrical and economical enquiries, however, the regression does not diverge very markedly from the linear form. In the cases of non-linear regression that I have hitherto had to deal with, I find that parabolæ of the 2nd or 3rd order will suffice as a rule to describe the deviation from linearity. If they did not, we could, of course, use curves of higher orders, but the difficulty referred to in the first section of this paper at once arises: we then need to use in the determination moments and product-moments of such high orders that the probable errors of the constants are so high as to render valueless their calculation from such statistical data as we can hope for in most actual inquiries. In the great bulk of investigations it is practically impossible to increase our random samples from 500 to 1,000 individuals up to 50,000 to 100,000. Nor in the great bulk of statistical cases is any such increase even desirable, for a fairly wide experience shows that 2nd and 3rd order parabolæ amply suffice to describe the skewness of the regression line. I shall accordingly classify skew correlation in the following manner:—

(*a.*) *Linear Regression:*

The mean of an x-array of y's, *i.e.*, y_{x_p}, is given by

$$y_{x_p}=a_0+a_1x_p \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxxviii.)}.$$

(*b.*) *Parabolic* Regression:*

The mean of an x-array of y's, *i.e.*, y_{x_p}, is given by

$$y_{x_p}=a_0+a_1x_p+a_2x_p^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xxxix.)}.$$

(*c.*) *Cubical* Regression:*

The mean of an x-array of y's, *i.e.*, y_{x_p}, is given by

$$y_{x_p}=a_0+a_1x_p+a_2x_p^2+a_3x_p^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xl.)}.$$

It is conceivable—in fact, from unpublished work already done, highly probable—that the theory of skew variation will give regression curves, not of the exact form involved in (xxxix.) or (xl.), but containing product terms in x and y. The most general equation to a regression curve may be taken to be of the type

$$y_{x_p}-\bar{y}=f(x_p-\bar{x}),$$

and what experience shows us is: that for the great bulk of vital phenomena it is sufficient to expand by MACLAURIN'S theorem and keep the first three or four terms. Indeed, in the large majority of cases, (xxxviii.) alone suffices. Hence, if (xxxix.) or (xl.) fit the data within the limits of random sampling, we are not injudiciously circumscribing future developments of the theory of skew correlation by casting our regression curves into the above forms. I shall deal first with the theory of cubical regression, for we can then obtain from this the conditions necessary for parabolic and linear regressions.

I must remind the reader, however, that the form of the regression line does not in any way limit the nature of the distribution of the array about its mean; the variability of an array, *i.e.*, the standard deviation of an array, having for its mean value $\sigma_y\sqrt{1-\eta^2}$, may or may not be the same for all arrays. If it is the same, or all arrays are *equally scattered* about their means, I shall speak of the system as a *homoscedastic* system, otherwise it is a *heteroscedastic* system. The Gauss-Laplacian correlation surface gives a homoscedastic linear system. Mr. YULE'S linear regression is not necessarily homoscedastic; it may, however, be homoscedastic without being normal, and then the scatter of each array is measured by $\sigma_y\sqrt{1-r^2}$. When a system is homoscedastic, but not linear, then $\sigma_{n_{x_p}}{}^2=\sigma_y{}^2(1-\eta^2)$, and consequently the χ_1 of (xxxv.) is equal to unity. $\chi_1=1$ is a necessary result of homoscedasticity.

Lastly, we want a word to express the idea of all the arrays having equal skewness,

* 'Parabolic' and 'cubical' are here used in the narrower sense of regression curves corresponding to ordinary parabolæ of the 2nd order and of the 3rd order respectively: in both cases the axis of the parabola being parallel to the axis of the y-character.

or being asymmetrical in an equal degree about their means. I shall express this by the term *homoclitic;* generally the arrays will not be equally asymmetrical round their means, and in this case we shall speak of them as *heteroclitic.* If there were no skewness in any of the arrays, then m_3 of (xxxvi.) would be zero for all of them. I term arrays of no skewness *isocurtic*, and skew arrays *allocurtic.* If we supposed that a curve of Type III. would sufficiently express the skewness of an array, we should have

$$\text{Sk.}=\tfrac{1}{2}m_3/\sigma_{n_{x_p}}{}^3,$$

and therefore from (xxxvi.)

$$\chi_2=\frac{2\text{S}\{n_{x_p}\sigma_{n_{x_p}}{}^3(\text{Sk.})(y_{x_p}-\bar{y})\}}{\text{N}\sigma_y{}^3(1-\eta^2)^{3/2}\sigma_{\text{M}}} \quad \ldots\ldots\ldots \text{(xli.)}.$$

For a homoscedastic system we have $\sigma_{n_{x_p}}=\sigma_y\sqrt{1-\eta^2}$, and therefore

$$\chi_2=\frac{2\text{S}\{n_{x_p}(\text{Sk.})(y_{x_p}-\bar{y})\}}{\text{N}\sigma_{\text{M}}},$$

and for a homoclitic system

$$\chi_2=\frac{2(\text{Sk.})\,\text{S}\{n_{x_p}\sigma_{n_{x_p}}{}^3(y_{x_p}-\bar{y})\}}{\text{N}\sigma_y{}^3(1-\eta^2)^{3/2}\sigma_{\text{M}}}.$$

For a homoclitic homoscedastic system, whether isocurtic or allocurtic,

$$\chi_2=\frac{2(\text{Sk.})\,\text{S}\{n_{x_p}(y_{x_p}-\bar{y})\}}{\text{N}\sigma_{\text{M}}}=0.$$

Thus χ_2 is to a certain extent a measure of both homoscedasticity and homoclisy. But as the correlation between $\sigma_{n_{x_p}}$ and $y_{x_p}-\bar{y}$ is in most cases extremely small, while the skewness of the array can well change its sign with arrays above or below the mean, we can fairly consider the smallness of χ_2 to be a measure of the approach to homoclisy. I am thus inclined to speak of χ_1-1 and χ_2 as measures of heteroscedasticity and heteroclisy. When they both vanish we have a homoscedastic homoclitic system. For such systems η, the correlation ratio, tells us effectively the scatter of any array, and as a rule all we want to know, in addition, is the form of the regression line.

(5.) *Cubical Regression.*

We have already used the following notation

$$\text{N}p_{qq'}=\text{S}\{n_{x_y}(x-\bar{x})^q(y-\bar{y})^{q'}\} \quad \ldots\ldots\ldots \text{(xlii.)}.$$

We shall shorten our formulæ if we write

$$r=p_{11}/(\sigma_x\sigma_y), \quad \epsilon=p_{21}/(\sigma_x{}^2\sigma_y), \quad \zeta=p_{31}/(\sigma_x{}^3\sigma_y), \quad \theta=p_{41}/(\sigma_x{}^4\sigma_y) \quad \text{(xliii.)}.$$

We have already used μ_q to denote p_{0q}, and we shall use ν_q for p_{q0}. Further, we write

$$\beta_1=\nu_3{}^2/\nu_2{}^3, \quad \beta_2=\nu_4/\nu_2{}^2, \quad \beta_3=\nu_5\nu_3/\nu_2{}^4, \quad \beta_4=\nu_6/\nu_2{}^3 \quad \ldots\ldots \text{(xliv.)}.$$

$\sqrt{\beta_1}=\nu_3/\sigma_x^3$ will be of the same sign as ν_3. These constants β have been previously used in the theory of skew variation.*

We shall further put

$$\bar{\epsilon}=\epsilon-r\sqrt{\beta_1},\quad \bar{\zeta}=\zeta-r\beta_2,\quad \bar{\theta}=\theta-r\beta_3/\sqrt{\beta_1} \quad . \quad . \quad . \quad . \quad \text{(xlv.)}.$$

The regularity of the forms $\bar{\epsilon}$, $\bar{\zeta}$, $\bar{\theta}$, is rather screened by the above notation, which is introduced for brevity; using the $p_{qq'}$ notation, we have

$$\bar{\epsilon}=\frac{p_{21}p_{20}-p_{11}p_{30}}{\sigma_x^4\sigma_y},\quad \bar{\zeta}=\frac{p_{31}p_{20}-p_{11}p_{40}}{\sigma_x^5\sigma_y},\quad \bar{\theta}=\frac{p_{41}p_{20}-p_{11}p_{50}}{\sigma_x^6\sigma_y} \quad . \quad . \quad \text{(xlvi.)},$$

whence the law of formation of these constants is easily seen.

The regression curve may now be conveniently put into the form

$$\frac{y_{x_p}-\bar{y}}{\sigma_y}=b_0+b_1\frac{x_p-\bar{x}}{\sigma_x}+b_2\left(\frac{x_p-\bar{x}}{\sigma_x}\right)^2+b_3\left(\frac{x_p-\bar{x}}{\sigma_x}\right)^3 \quad . \quad . \quad . \quad . \quad \text{(xlvii.)}.$$

Or, multiplying by n_{x_p} and summing for all arrays,

$$0=\mathrm{N}b_0+b_2\mathrm{N}+b_3\mathrm{N}\sqrt{\beta_1},$$

the sign of $\sqrt{\beta_1}$ being always that of the 3rd moment. Hence, measuring from the means of the two characters, *i.e.*, $\mathrm{X}_p=x_p-\bar{x}$, $\mathrm{Y}_{x_p}=y_{x_p}-\bar{y}$, we may re-write (xlvii.)

$$\mathrm{Y}_{x_p}/\sigma_y=b_1(\mathrm{X}_p/\sigma_x)+b_2\{(\mathrm{X}_p/\sigma_x)^2-1\}+b_3\{(\mathrm{X}_p/\sigma_x)^3-\sqrt{\beta_1}\} \quad . \quad . \quad \text{(xlviii.)}.$$

Now multiply by $n_{x_p}\mathrm{X}_p/\sigma_x$ and sum for all arrays, remembering that

$$\mathrm{N}r\sigma_x\sigma_y=\mathrm{S}(n_{x_y}\mathrm{XY})=\mathrm{S}(n_{x_y}\mathrm{X}_p\mathrm{Y}_{x_p}),$$

we find

$$r=b_1+b_2\sqrt{\beta_1}+b_3\beta_2.$$

This enables us to get rid of b_1 and write (xlviii.)

$$\mathrm{Y}_{x_p}/\sigma_y=r\mathrm{X}_p/\sigma_x+b_2\{(\mathrm{X}_p/\sigma_x)^2-\sqrt{\beta_1}(\mathrm{X}_p/\sigma_x)-1\}$$
$$+b_3\{(\mathrm{X}_p/\sigma_x)^3-\beta_2(\mathrm{X}_p/\sigma_x)-\sqrt{\beta_1}\} \quad . \quad . \quad . \quad \text{(xlix)}.$$

Now multiply by $n_{x_p}(\mathrm{X}_p/\sigma_x)^2$ and sum for all arrays. We have

$$\epsilon=r\sqrt{\beta_1}+b_2(\beta_2-\beta_1-1)+b_3(\beta_3/\sqrt{\beta_1}-\beta_2\sqrt{\beta_1}-\sqrt{\beta_1}),$$

or

$$\bar{\epsilon}=b_2\phi_2+b_3\phi_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(l.)},$$

where

$$\left.\begin{aligned}\phi_2&=\beta_2-\beta_1-1\\ \phi_3&=(\beta_3-\beta_1\beta_2-\beta_1)/\sqrt{\beta_1}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(li.)}.$$

* 'Phil. Trans.,' A, vol. 186, p. 368, and A, vol. 198, p. 278.

Eliminating b_2, we can write (xlix.)

$$Y_{x_p}/\sigma_y = r(X_p/\sigma_x) + \frac{\bar{\epsilon}}{\phi_2}\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\}$$
$$+ b_3\left[(X_p/\sigma_x)^3 - \beta_2(X_p/\sigma_x) - \sqrt{\beta_1} - \frac{\phi_3}{\phi_2}\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\}\right] . \quad \text{(lii.)}.$$

Now multiply by $n_{x_p}(X_p/\sigma_x)^3$ and sum for all arrays; we find

$$\zeta = r\beta_2 + \frac{\bar{\epsilon}}{\phi_2}\phi_3 + b_3(\phi_4 - \phi_3^2/\phi_2),$$

or

$$(\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)/(\phi_2\phi_4 - \phi_3^2) = b_3 \quad \text{(liii.)},$$

where

$$\phi_4 = \beta_4 - \beta_2^2 - \beta_1 \quad \text{(liv.)}.$$

It follows from (l.) that

$$b_2 = (\bar{\epsilon}\phi_4 - \bar{\zeta}\phi_3)/(\phi_2\phi_4 - \phi_3^2) \quad \text{(lv.)}.$$

We can thus write the cubic regression curve in either of the forms*

* The method is perfectly easy of extension, if we choose to use higher products and moments, to a regression curve of any order, *e.g.*,

$$Y_{x_p}/\sigma_y = b_0 + b_1(X_p/\sigma_x) + b_2(X_p/\sigma_x)^2 + \ldots + b_n(X_p/\sigma_x)^n + \ldots$$

For let: $N\epsilon_{q1} = S(n_{xy}Y_{x_r}X_p{}^q)/(\sigma_x{}^q\sigma_y)$, and $\gamma_s = \nu_s/\sigma_x{}^s = S(n_{x_r}X_p{}^s)/(N\sigma_x{}^s)$,

we have:

$$0 = b_0 + 0\times b_1 + b_2 + \gamma_3 b_3 + \ldots + \gamma_n b_n + \ldots$$
$$\epsilon_{11} = 0\times b_0 + b_1 + \gamma_3 b_2 + \gamma_4 b_3 + \ldots + \gamma_{n+1} b_n + \ldots$$
$$\epsilon_{21} = b_0 + \gamma_3 b_1 + \gamma_4 b_2 + \gamma_5 b_3 + \ldots + \gamma_{n+2} b_n + \ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\epsilon_{p1} = \gamma_p b_0 + \gamma_{p+1} b_1 + \gamma_{p+2} b_2 + \gamma_{p+3} b_3 + \ldots + \gamma_{n+p} b_n + \ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Hence writing ϵ_{01} for 0, $\gamma_0 = 1$, $\gamma_1 = 0$, $\gamma_2 = 1$, we have

$$b_n = (\epsilon_{01}\Delta_{0n} + \epsilon_{11}\Delta_{1n} + \epsilon_{21}\Delta_{2n} + \ldots + \epsilon_{p1}\Delta_{pn} + \ldots)/\Delta,$$

where $\Delta =$

$$\begin{vmatrix} \gamma_0, & \gamma_1, & \gamma_2, & \gamma_3, & \ldots & \gamma_n, & \ldots \\ \gamma_1, & \gamma_2, & \gamma_3, & \gamma_4, & \ldots & \gamma_{n+1}, & \ldots \\ \gamma_2, & \gamma_3, & \gamma_4, & \gamma_5, & \ldots & \gamma_{n+2}, & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \gamma_p, & \gamma_{p+1}, & \gamma_{p+2}, & \gamma_{p+3}, & \ldots & \gamma_{p+n}, & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{vmatrix}$$

and Δ_{qn} is the minor of the constituent in the $(q+1)^{\text{th}}$ row and $(n+1)^{\text{th}}$ column. As we have already noted, however, solutions involving anything beyond γ_6 are hardly likely to be of practical value.

The value above for b_n is the type equation given by the method of least squares, when we strike the best fitting curve to all the entries in the correlation table. I have already pointed out that the method of moments becomes identical with that of least squares, when we fit parabolæ of any order ('Biometrika,' vol. I., p. 271). The retention of the method of moments, however, enables us, without abrupt change of method, to introduce the needful η, and to grasp at once the application of the proper SHEPPARD'S corrections. The extension of the method of least squares to *continua* in space has not yet, as far as I am aware, been fully considered.

$$Y_{x_p}/\sigma_y = r(X_p/\sigma_x) + \frac{\bar{\epsilon}}{\phi_2}\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\}$$
$$+ \frac{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}{\phi_2\phi_4 - \phi_3^2}\left[(X_p/\sigma_x)^3 - \beta_2(X_p/\sigma_x) - \sqrt{\beta_1} - \frac{\phi_3}{\phi_2}\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\}\right] . \quad \text{(lvi.)}.$$

or

$$Y_{x_p}/\sigma_y = r(X_p/\sigma_x) + \frac{\bar{\epsilon}\phi_4 - \bar{\zeta}\phi_3}{\phi_2\phi_4 - \phi_3^2}\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\}$$
$$+ \frac{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}{\phi_2\phi_4 - \phi_3^2}\{(X_p/\sigma_x)^3 - \beta_2(X_p/\sigma_x) - \sqrt{\beta_1}\} \quad . \quad . \quad \text{(lvi.) } \textit{bis.}$$

The former arrangement of the solution, while it is apparently more cumbersome, is, perhaps, the better, for it gives us at once the measure of the deviation from parabolic or 2nd order regression, *i.e.*, the approach of $\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3$ to zero. In the case of normal correlation both $\bar{\epsilon}$ and $\bar{\zeta}$ vanish, and neglecting higher terms the condition for linear regression is that $\bar{\epsilon}=0$, and $\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3 = 0$, or, again, $\bar{\epsilon}$ and $\bar{\zeta}=0$. For material in which the x-variability is isocurtic, $\beta_1=\beta_3=\phi_3=0$, and the regression curve takes the simple form

$$Y_{x_p}/\sigma_y = r(X_p/\sigma_x) + \frac{\bar{\epsilon}}{\phi_2}\{(X_p/\sigma_x)^2 - 1\} + \frac{\bar{\zeta}}{\phi_4}\{(X_p/\sigma_x)^3 - \beta_2(X_p/\sigma_x)\} \quad . \quad \text{(lvi.) } \textit{ter.}$$

We now turn to express these relations in terms of the correlation ratio η. Multiply (lvi.) by $n_{x_p}Y_{x_p}/\sigma_y$, and sum for all arrays, we obtain

$$\eta^2 = r^2 + \frac{\bar{\epsilon}}{\phi_2}(\epsilon - \sqrt{\beta_1}r) + \frac{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}{\phi_2\phi_4 - \phi_3^2}\left\{\zeta - \beta_2 r - \frac{\phi_3}{\phi_2}(\epsilon - \sqrt{\beta_1}r)\right\},$$

whence results

$$\phi_2(\eta^2 - r^2) - \epsilon^2 = (\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)^2/(\phi_2\phi_4 - \phi_3^2) \quad . \quad . \quad . \quad . \quad . \quad \text{(lvii.)}.$$

(lvii.) is a necessary condition of cubical regression.

It is of course not a sufficient condition, as we ought to show that b_4, b_5, &c., all vanish, and thus any number of conditions may be found. For example, multiply by $n_{x_p}X_p^4/\sigma_x^4$ and sum for all arrays, then

$$\theta = \frac{\bar{\epsilon}\phi_4 - \bar{\zeta}\phi_3}{\phi_2\phi_4 - \phi_3^2}(\beta_4 - \beta_3 - \beta_2) + \frac{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}{\phi_2\phi_4 - \phi_3^2}\frac{\beta_5 - \beta_2\beta_3 - \beta_1\beta_2}{\sqrt{\beta_1}} \quad . \quad . \quad . \quad \text{(lviii.)}$$

is also a necessary condition. Here $\beta_5 = \nu_7\nu_3/\sigma_x^{10}$. But the high as well as complicated value of the probable errors of such expressions renders it idle to consider them in practice.

Substituting (lvii.) in (lvi.) we have :

$$\mathrm{Y}_{x_p}/\sigma_y = r\,(\mathrm{X}_p/\sigma_x) + \frac{\bar{\epsilon}}{\phi_2}\{\chi_p/\sigma_x)^2 - \sqrt{\beta_1}\,(\chi_p/\sigma_x) - 1\}$$
$$\pm \sqrt{\frac{\phi_2(\eta^2 - r^2) - \bar{\epsilon}^2}{\phi_2\phi_4 - \phi_3^2}}\Big[(\mathrm{X}_p/\sigma_x)^3 - \beta_2\,(\chi_p/\sigma_x) - \sqrt{\beta_1}$$
$$- \frac{\phi_3}{\phi_2}\{(\mathrm{X}_p/\sigma_x)^2 - \sqrt{\beta_1}(\mathrm{X}_p/\sigma_x) - 1\}\Big] \quad . \quad \text{(lix.)}.$$

Which sign is to be given to the root will often be visible on inspection of the observations. Otherwise the sign of the root must be the same as that of

$$\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3.$$

(lix.) will save the calculation of $\bar{\zeta}$ if the root-sign can be found by inspection.

Finally there is a third form into which we may put the cubic. Eliminate $\phi_2\phi_4 - \phi_3^2$ from (lix.) by aid of (lvii.) and it becomes

$$\mathrm{Y}_{x_p}/\sigma_y = r\,(\mathrm{X}_p/\sigma_x) + \frac{\bar{\epsilon}\bar{\zeta} - \phi_3\,(\eta^2 - r^2)}{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}\{(\mathrm{X}_p/\sigma_x)^2 - \sqrt{\beta_1}\,(\mathrm{X}_p/\sigma_x) - 1\}$$
$$+ \frac{\phi_2(\eta^2 - r^2) - \bar{\epsilon}^2}{\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3}\{(\mathrm{X}_p/\sigma_x)^3 - \beta_2\,(\mathrm{X}_p/\sigma_x) - \sqrt{\beta_1}\} \quad . \quad . \quad \text{(lx.)}.$$

At first sight this might appear to be the best form of the cubic, because it does not involve the 6th moment of the variable x. But this is very far from being the case in actual practice. The reason is simply this, $\bar{\epsilon}$, $\bar{\zeta}$ and $\eta^2 - r^2$ are in most cases very small—they vanish in normal correlation—relatively to ϕ_2 and ϕ_4. Hence both numerators and denominators of the coefficients of the square and cubic terms are the ratio of small quantities, and accordingly subject to large probable errors. For this reason (lx.) was found in actual practice to be of no service. Of the other two forms (lvii.) and (lix.), which neither suffer from this defect, $\phi_2\phi_4 - \phi_3^2$ being always large relative to the numerators, (lix.) while involving a 6th moment does not involve a 4th product, $\bar{\zeta}$, and experience shows that the former is on the whole easier to determine and more exact than the former. Hence (lix.) seems the preferable form, even if it be needful in certain cases to determine $\bar{\zeta}$ in order to fix the sign of the radical. The cubic regression curve thus demands a knowledge of the correlation ratio η, of the "cubic product" $\bar{\epsilon}$ and the sign by inspection or calculation of $\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3$. Besides this, we require the first six moments of the independent variable x. Of course if the regression of x on y be required, as well as that of y on x, the second correlation ratio and cubic product as well as the first six moments of y must be found. It is rare, however, that both regression curves are needed for a single enquiry.

As to the general form of (lix.), we note that there will always be a real point of inflexion given by

$$\mathrm{X}_p/\sigma_x = \tfrac{1}{3}\,(b_3\phi_3 - \bar{\epsilon})/(b_3\phi_2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxi.)},$$

where

$$b_3=\pm\sqrt{\{\phi_2(\eta^2-r^2)-\bar{\epsilon}^2\}/(\phi_2\phi_4-\phi_3^2)},$$

and further that there may be two points of horizontality given by a certain quadratic. Thus, in general, the regression line will tend to be part of an **S**-shaped curve. The horizontal points may be imaginary, or, if real, either they or the point of inflexion may be far beyond the portion of the curve which crosses the observed field of frequency. If we consider, however, the slope of the regression curve to measure the regression in the neighbourhood of any point, we note that the regression is a maximum at the point given by (lxi.), and grows smaller and smaller towards the two points of horizontality, *i.e.*, points of complete local independence of the two characters. These are not unfamiliar features in certain practical cases of skew correlation,* and accordingly the cubic regression curve provides us with a ready means of describing regression phenomena, which cannot be dealt with by the simple line or the parabola.

It may of course be suggested that a quartic or quintic curve would give a better result than a cubic. The answer to this is: Possibly, but the high moments and products required render it impossible to deal even superficially with the probable errors of the constants involved. The calculation of the probable error of η is a sufficiently stiff task in the general case. To test the probable error of a condition like (lvii.), to say nothing of one like (lviii.), would involve an immense amount of work, since we should want the correlation of errors in η, $\bar{\epsilon}$, ζ, and $\bar{\theta}$. Speaking with some experience of practical statistical possibilities, I think, the tendency to use very high moments or product-moments must be curtailed to the minimum of actual needs. We cannot deny the existence of skew variation, nor of the sensible curvature of regression lines. We must admit their existence as the result of statistical experience. This existence involves a great widening of the old frequency notions and the need for a new means of description. But we must remember that statistics are essentially a practical study, the art of describing by a few numerical constants observational experience, and we must curtail at every turn the desire to run riot in mathematical formulæ, which cannot be generally applied in actual practice.† Still I propose later in this paper to deal with the general formulæ for quartic regression.

(6.) *Parabolic Regression.*

For a parabolic system b_3 must vanish, or nearly vanish. Hence we have from (liii.) and (lvii.).

$$\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxii.)},$$

$$\phi_2(\eta^2-r^2)-\bar{\epsilon}^2=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxiii.)}.$$

* Compare for example the regression line of age of mean age of bridegroom for actual age of bride, which gives a typical **S**-shaped curve. See 'Biometrika,' vol. II., p. 20.

† These remarks have special reference to the points dealt with on p. 6.

From these conditions we find

$$b_2=\bar{\epsilon}/\phi_2=\pm\sqrt{(\eta^2-r^2)/\phi_2}.$$

These give for the form of the parabolic regression curve

$$Y_{x_p}/\sigma_y=r\,(X_p/\sigma_x)+\frac{\bar{\epsilon}}{\phi_2}\{(X_p/\sigma_x)^2-\sqrt{\beta_1}\,(X_p/\sigma_x)-1\} \quad . \quad . \quad . \quad \text{(lxiv.)},$$

or

$$Y_{x_p}/\sigma_y=r\,(Y_{px}/\sigma_x)\pm\sqrt{\frac{\eta^2-r^2}{\phi_2}}\{(X_p/\sigma_x)^2-\sqrt{\beta_1}\,(X_p/\sigma_x)-1\} \quad . \quad . \quad \text{(lxv.)}.$$

The latter form, besides the correlation coefficient and correlation ratio, requires only a knowledge of the skew variation constants β_1 and β_2, and is therefore very easy to determine. Except for very nearly linear regression, there can be no doubt as to the sign of $\sqrt{\eta^2-r^2}$, as we can tell at once whether the parabola ought to be concave or convex to the x-axis. In other cases the sign of $\sqrt{\eta^2-r^2}$ must be taken to coincide with that of $\bar{\epsilon}$, which must therefore be found. It will then be as easy to use (lxiv.) as (lxv.), although probably η and r can be found with less error than $\bar{\epsilon}$.

It is thus quite easy to allow for such curvature of the regression line as can be expressed by a parabola of the 2nd order of the type considered.

We notice at once that the regression curve does not pass through the mean of the two characters. Or, an individual with the mean of one character will most probably not have the mean of a second character. This is a rather important result, which follows at once for nearly all types of skew correlation.

It will be seen, for example, that QUETELET's "mean man," defended by Professor EDGEWORTH as theoretically justifiable, depends entirely on human characters giving linear regression curves. Such linear curves are certainly given by many pairs of characters, *e.g.*, cranial and body measurements, but there are certainly other characters for which regression ceases to be sensibly linear, and the conception of the "mean man" in this case fails. For example, if age be considered as a character, then the regression is certainly not linear, and the individual of mean age will not necessarily have either the mean physical or psychical characters. This seems of some importance for the general conception of "type," if by type we denote the mean, for probably there are other characters than age for which regression is skew.

The regression, *i.e.*, dY_{x_p}/dX_p will be zero, for a point $X_{(Y\,max.)}$ for which

$$\frac{X_{(Y\,max.)}}{\sigma_x}=\tfrac{1}{2}\left\{\sqrt{\beta_1}-r\sqrt{\frac{\beta_2-\beta_1-1}{\eta^2-r^2}}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxvi.)}$$

the sign of the root being determined as before. Clearly, therefore, unless r be very small, or η^2 diverges very sensibly from r^2, this point of zero regression may correspond

to a very large abscissa, and in some cases will lie entirely outside the range of observable frequency.

The parabola of regression cuts the line of regression, *i.e.*, the line of best fit to the series of regression points, or to the means of the x-arrays, in two points determined by the quadratic equation

$$\left(\frac{X_p}{\sigma_x}\right)^2-\sqrt{\beta_1}\frac{X_p}{\sigma_x}-1=0,$$

or

$$\frac{X_p}{\sigma_x}=\tfrac{1}{2}\{\sqrt{\beta_1}\pm\sqrt{\beta_1+4}\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxvii.)}.$$

These points are always real, and correspond, if regression be truly parabolic, to the same values of the x-character, whatever be the y-character of which we are considering the correlation. In the case of normal variation of the x-character only, these are the points of inflexion of the x-distribution.

(7.) *Linear Regression.*

In this case it is necessary that both b_2 and b_3 vanish within the limits of random sampling, and, although these are not theoretically sufficient—for a whole series of relations between the higher product-moments could be written down*—they are for practical purposes sufficient.

Hence we have the following conditions for linear regression :—

$$\eta^2=r^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxviii.)},$$

or, the coefficient of correlation, without regard to sign, should be equal to the correlation ratio. Further $\bar{\epsilon}$ should be zero, or

$$p_{21}p_{20}-p_{11}p_{30}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxix.)}.$$

The theory of linear regression is so familiar that it need not be further discussed here. In the actual practice of statistics, the determination of the means of the x-arrays and the drawing of the regression line will often suffice to show the fairly trained eye whether the deviations from it are random or not. If they are not random, then we must proceed to the determination of η and of the higher product-moments.

The following are numerical examples of skew correlation, selected to illustrate the theory developed above.

* For example, it is necessary in most cases that $\bar{\zeta}$ should vanish. In the instance of that very special case of linear regression, the Gauss-Laplacian normal frequency, it is easy to show that the constants $\bar{\epsilon}$, $\bar{\zeta}$ both vanish as well as $\eta^2=r^2$.

Statistical Illustrations.

(8.) *Illustration* A.—*On the Skew Correlation between Number of Branches to the Whorl and Position of the Whorl on the Spray in the case of* Asperula odorata.

In this case the material was collected in a lane near Horsham, Sussex, at Whitsuntide, 1903, by Miss M. Radford. There were 150 independent sprays, the woodruff had just flowered, and the whorls were counted from the flower *downwards*. Being early in the season, the maximum number of whorls was five, and, in some cases, not even as many were available. The material was counted and tabled by the author, and the results are exhibited in the table below :—

Table I.—Correlation of Whorl-Branches and Position of Whorl.

	x.	Whorl.	Number of branches in whorl.					n_p.	y_{x_p}.	σ_{n_p}.	m_2.	m_3.
			4.	5.	6.	7.	8.					
Position of whorl.	x_1	First . .	—	3	66	42	39	150	6·7800	·8553	·7316	·1535
	x_2	Second .	—	3	61	47	39	150	6·8133	·8437	·7117	·0985
	x_3	Third . .	—	6	60	40	44	150	6·8133	·9047	·8185	·0383
	x_4	Fourth .	1	12	68	39	22	142	6·4859	·8780	·7709	·1347
	x_5	Fifth . .	1	13	53	10	10	87	6·1724	·8605	·7404	·4049
Totals. . . .			2	37	308	178	154	679	6·6554	—	—	—

We require the regression curve giving the probable number of branches for a given whorl.

Dealing first with the skew variation in position, a purely arbitrary system depending solely on the number of whorls dealt with in each position, we find, not using Sheppard's correction,*

$$\text{Mean}=2\cdot802{,}651, \quad \nu_2=1\cdot787{,}268, \quad \nu_5=2\cdot799{,}638,$$
$$\sigma_x=1\cdot336{,}887, \quad \nu_3=\cdot311{,}783, \quad \nu_6=22\cdot678{,}308.$$
$$\nu_4=5\cdot841{,}682.$$

Hence we determine

$$\beta_1=\cdot017{,}027, \quad \phi_2=\cdot811{,}740,$$
$$\beta_2=1\cdot828{,}767, \quad \phi_3=\cdot286{,}465.$$
$$\beta_3=\cdot085{,}545, \quad \phi_4=\cdot610{,}879,$$
$$\beta_4=3\cdot972{,}295, \text{ and } \sqrt{\beta_1}=+\cdot130{,}487.$$

* The numbers are tabulated to six places, because we cannot be sure that the final calculations are for the data true to two places, which is all we finally retain unless this is done. Any number of figures can really be retained with perfect ease when the work is done on a calculator.

We now turn to the skew variation in the number of branches to the whorl, and get the following constants :—

$$\text{Mean}=6\cdot655{,}375, \qquad \mu_2=\cdot806{,}124,$$
$$\sigma_y=\cdot897{,}842, \qquad \mu_3=\cdot132{,}090,$$
$$\mu_4=1\cdot138{,}410.$$

The values of y_{x_p}, m_2, and m_3 are given in table above. Using them we find

$$\sigma_M=\cdot224{,}377, \qquad \eta=\cdot249{,}911, \qquad \sigma_{x_y}=\sigma_y\sqrt{1-\eta^2}=\cdot869{,}355,$$
$$\lambda_2=\sigma_M{}^2=\cdot050{,}345, \qquad \lambda_4=\cdot007{,}474, \qquad \chi_1=\cdot990{,}862, \quad \chi_2=-\cdot059{,}851.$$

These give by (xxxiii.), showing the numerical contribution of each term,

$$\Sigma_\eta{}^2=\frac{1}{N}\{\cdot878{,}991-\cdot010{,}323-\cdot000{,}888-\cdot007{,}231+\cdot013{,}578\},$$

or the probable error of $\eta=\cdot0242$.

Had we calculated the probable error of η from (xxxiv.), we should have found for its value ·0243. It is clear that for this special case the simple formula (xxxiv.) is amply sufficient, the small terms almost cancelling.

We see that χ_1 is almost unity, and the graph of σ_{x_y}/σ_y shows indeed that the system is sensibly homoscedastic. χ_2 is small, but a glance at the graph of the clitic curve on Diagram I. shows that we can hardly treat the system as homoclitic, the changes in the skewness forming a fairly uniform curve.*

For practical purposes, we may treat the variability of the number of branches in any array as sufficiently closely given by $\sigma_y\sqrt{1-\eta^2}$.

We now turn to the product-moments† and find

$$p_{11}=-\cdot249{,}160, \qquad p_{31}=-\cdot896{,}415,$$
$$p_{21}=-\cdot236{,}289, \qquad p_{41}=-1\cdot210{,}225.$$

* Throughout these illustrations the clitic curve is plotted by calculating the skewness of the arrays from $\frac{1}{2}m_3/(m_2)^{3/2}$. See p. 23.

† In calculating these products referred to the centroid from those referred to any axes, generally corresponding to whole numbers in the table, the following reduction formulæ will be found useful We take $N\Pi_{qq'}=S(n_{xy}\,x'^q y'^{q'})$, x' and y' being measured from any axes, further, $\bar{x}'$, $\bar{y}'$ are the distances of the means from these axes, and ν_2, ν_3, ν_4 the moments of the x-character about its mean as tabled above.

$$p_{11}=\Pi_{11}-\bar{x}'\Pi_{01}, \qquad p_{21}=\Pi_{21}-2\bar{x}'\Pi_{11}+\bar{x}'^2\Pi_{01}-\bar{y}'\nu_2,$$
$$p_{31}=\Pi_{31}-3\bar{x}'\Pi_{21}+3\bar{x}'^2\Pi_{11}-\bar{x}'^3\Pi_{01}-\bar{y}'\nu_3,$$
$$p_{41}=\Pi_{41}-4\bar{x}'\Pi_{31}+6\bar{x}'^2\Pi_{21}+4\bar{x}'^3\Pi_{11}+\bar{x}'^4\Pi_{01}-\bar{y}'\nu_4.$$

The p's should be further corrected for grouping by SHEPPARD'S corrections (given on my p. 36), provided there be high contact at the contour of the surface of frequency. SHEPPARD'S corrections have not in this

These lead to

$$r=-\cdot207,579, \quad \bar{\epsilon}=-\cdot120,164, \quad \bar{\zeta}=-\cdot038,241, \quad \bar{\theta}=-\cdot285,890.$$

Thus all the constants are determined.

We find

$$\eta^2-r^2=\cdot019,367,$$
$$\phi_2(\eta^2-r^2)-\bar{\epsilon}^2=\cdot001,281,$$
$$\phi_2(\eta^2-r^2)-\bar{\epsilon}^2-(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2/(\phi_2\phi_4-\phi_3^2)=\cdot000,276.$$

These should be respectively zero for linear, parabolic, and cubical regressions. It will be seen that they are satisfied with increasing closeness; we might well be satisfied even with the parabolic regression curve. The following are the regression curves determined, y_{x_p} being the actual number of branches in the whorl ($=6\cdot655,375+Y_{x_p}$), and x_p the actual position of the whorl :—

(*a.*) *Straight line:*

$$y_{x_p}=7\cdot046,087-\cdot139,408\,x_p.$$

(*b.*) *Parabola* from (lxv.):

$$y_{x_p}=6\cdot794,052-\cdot125,872\,x_p-\cdot077,592\,x_p^2;$$

or,

$$y_{x_p}=6\cdot853,561-\cdot077,592\,(x_p-1\cdot991,535)^2.$$

This clearly gives a maximum number of branches, 6·8536 corresponding to $x_p=1\cdot9915$, a value within the limits of observation.

(*c.*) *Cubic* from (lix.):

$$y_{x_p}=6\cdot799,399-\cdot192,439\,X_p-\cdot084,230\,X_p^2+\cdot020,915\,X_p^3.$$

Here X_p is measured from the mean position $=x_p-2\cdot802,651$, and y_{x_p} is, as before, the total number of branches for the given position.

Condition (lvii.) is so closely satisfied that we shall here get sensibly as good results from (lix.) as from (lvi.).

In the table below and in the curves of Diagram I. the values of the mean of the arrays, as found from line, parabola, and cubic, are given and compared with observation.

case been used, as this condition is not fulfilled. The axes x', y' actually taken for woodruff were those through the third whorl and through six branches.

An obvious warning about the signs of the sums of the products may be given which may save computators some trouble. The axes being taken positive, as in the accompanying figure, then the sums of the products for Π_{11} and Π_{31} are positive in the 1st and 3rd, negative in the 2nd and 4th quadrants. For Π_{21} and Π_{41} they are positive in the 1st and 4th quadrants and negative in the 2nd and 3rd quadrants. In the figure the axes are taken so as to suit the x and y-directions of the table on p. 31. Care must, of course, be paid to this point. The products may also be found from the y_{x_p}'s in the manner indicated on p. 35, footnote. They were thus verified in this case.

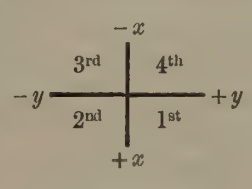


TABLE II.—Mean Branches to each Whorl.

$x_p =$	0.	1.	2.	3.	4.	5.	6.
y_{x_p} from line	[7·046]	6·907	6·767	6·628	6·488	6·349	6·210
y_{x_p} ,, parabola . . .	[6·546]	6·777	6·854	6·775	6·541	6·151	5·607
y_{x_p} ,, cubic	[6·117]	6·750	6·889	6·758	6·443	6·192	6·007
Observed	—	6·780	6·813	6·813	6·486	6·172	1

I think we may safely say that in the relationship of branches to position of the whorl in woodruff we have a case of homoscedastic correlation, which is effectively described by a parabolic regression curve. Thus, in a case of this kind, it is only needful, besides the moments up to the fourth of the x-character, to find the correlation coefficient r and the correlation ratio η.

(9.) *Illustration* B.—*On the Correlation between Age and Head Height in Girls.*

The data for this are taken from my School Measurement series, and involve the auricular heights of 2272 girls between the ages of 3 and 22. There was considerable paucity of material at the extreme ends of the range, and accordingly as our correlation curves are all obtained by weighting the observations, we can hardly expect good fits near 3 or 22 years of age. The actual correlation table is given as Table III. SHEPPARD'S corrections were applied throughout, and the unit of height is 2 millims.

In the first place the means, standard deviations, and 3rd moments of all the arrays of heights for different years of age were determined. These are given at the foot of Table III., but in actually calculating the constants more places of decimals were used. Then the first six moments of the frequency of the ages were found and the first four moments of the height frequencies. These are the x and y-frequencies. They give us :—

TABLE III.—Correlation between Age and Auricular Height of Head in Girls.

To face page 34.

Height of head.	Age. 3–4.	4–5.	5–6.	6–7.	7–8.	8–9.	9–10.	10–11.	11–12.	12–13.	13–14.	14–15.	15–16.	16–17.	17–18.	18–19.	19–20.	20–21.	21–22.	22–23.	Totals.	Height of head.
millims.																						millims.
102·25–104·25	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	102·25–104·25
104·25–106·25	—	—	—	2	—	1	1	1	2	1	2	—	—	—	—	—	—	—	—	—	10	104·25–106·25
106·25–108·25	—	—	1	—	1	—	1	—	4	—	2	—	1	—	—	—	—	—	—	—	10	106·25–108·25
108·25–110·25	—	—	—	1	5	2	1	4	2	2	4	1	3	1	—	—	1	—	—	—	27	108·25–110·25
110·25–112·25	—	—	1	3	1	5	12	3	6	5	3	9	2	4		1	—	—	—	—	56	110·25–112·25
112·25–114·25	—	—	1	—	4	3	10	8	6	9	4	3	5	5	1	—	—	—	—	—	59	112·25–114·25
114·25–116·25	1	—	3	4	7	8	15	14	11	16	10	7	6	8	2	2	—	1	—	—	115	114·25–116·25
116·25–118·25	—	2	2	9	9	7	10	23	15	18	13	9	11	6	4	3	—	—	1	—	142	116·25–118·25
118·25–120·25	—	2	2	4	13	22	24	25	37	44	23	11	19	6	6	3	2	1	—	—	244	118·25–120·25
120·25–122·25	—	2	3	6	9	19	25	29	34	41	32	21	15	13	9	4	—	—	3	—	265	120·25–122·25
122·25–124·25	—	—	3	3	7	17	23	34	38	33	21	22	18	25	9	4	1	2	—	1	261	122·25–124·25
124·25–126·25	—	—	—	1	6	19	18	33	29	40	32	23	26	14	12	10	—	1	1	—	265	124·25–126·25
126·25–128·25	—	—	1	6	9	10	8	21	27	27	32	20	18	16	13	9	1	—	—	1	219	126·25–128·25
128·25–130·25	—	—	—	—	—	6	9	17	16	20	39	25	29	16	11	7	—	1	1	—	197	128·25–130·25
130·25–132·25	—	—	—	1	3	5	5	7	13	17	17	15	18	12	6	6	4	1	1	—	131	130·25–132·25
132·25–134·25	—	—	—	—	1	—	7	8	10	13	8	5	16	7	7	6	—	—	—	—	88	132·25–134·25
134·25–136·25	—	—	—	—	1	1	3	4	4	9	11	13	9	11	8	2	1	—	—	—	77	134·25–136·25
136·25–138·25	—	—	—	—	—	—	3	2	2	10	4	5	14	6	3	2	1	—	—	—	52	136·25–138·25
138·25–140·25	—	—	—	—	—	—	—	2	3	3	2	2	2	4	2	—	—	—	—	—	20	138·25–140·25
140·25–142·25	—	—	—	—	—	—	1	—	2	1	2	4	2	2	—	1	1	—	—	—	16	140·25–142·25
142·25–144·25	—	—	—	—	—	—	1	—	—	—	2	3	—	4	—	1	—	—	—	—	11	142·25–144·25
144·25–146·25	—	—	—	—	—	—	—	—	—	—	—	—	—	2	1	—	—	—	1	—	4	144·25–146·25
146·25–148·25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1	146·25–148·25
Totals	1	7	18	40	76	125	177	235	261	309	263	198	214	162	95	61	13	7	8	2	2272	Totals.
Means in 1-millim. units	115·2500	116·9643	117·4722	119·1000	120·3026	121·6340	121·7246	122·8160	123·1427	123·8908	124·8622	125·7146	126·1565	126·5340	126·9132	127·0205	129·5577	123·8214	126·5000	125·2500	124·0467	Means in 1-millim. units.
Standard deviation in 2-millim. units	0	2·8853	2·9276	2·9641	2·9882	2·6366	3·3877	2·9653	3·2089	3·2061	3·3589	3·5865	3·4663	3·8696	3·1679	3·1235	4·8406	2·5311	4·1414	0·9574	3·4541	Standard deviation in 2-millim. units.
Third moments in 2-millim. units	0	− 42·822	− 18·108	− 7·679	+ 1·782	− 6·171	+ 15·893	+ 2·330	+ 0·238	+ 8·219	− 7·286	+ 3·015	− 9·615	+ 9·379	+ 2·991	+ 0·070	− 29·164	− 2·729	+ 85·816	0	+ 5·206	Third moments in 2-millim. units.

Height Constants.			*Age Constants.*		
Mean height =	124·0467 millims.		Mean age =	12·7007	
			σ_x=	3·064,819	in year units.
σ_y=	3·454,125	in 2 millim. units.	ν_2=	9·393,110	
μ_2=	11·930,977		ν_3=	1·051,882	
μ_3=	5·206,247		ν_4=	239·157,055	
μ_4=	438·639,633		ν_5=	104·298,702	
			ν_6=	9536·265,059	
β'_1=	·015,960,		β_1=	·001,335,	
β'_2=	3·081,454,		β_2=	2·710,593,	
			β_3=	·014,093,	
Further			β_4=	11·506,681,	
Σ_M=	2·093,366 millims.		$\sqrt{\beta_1}$=+	·036,538,	
λ_2=	4·382,181	in 1 millim. units.	ϕ_2=	1·709,258,	
λ_4=	62·399,135		ϕ_3=	·250,123.	
Hence					
$(\lambda_4-3\lambda_2^2)/(4\lambda_2^2)$=	·062,340,		ϕ_4=	4·158,032.	

In the next place the products were worked out and referred to the means with the following results :—*

p_{11}=	3·113,712,	whence r=	·294,128,
p_{21}= −	1·957,022,	$\bar{\epsilon}$= −	·071,065,
p_{31}=	74·447,616,	$\bar{\zeta}$= −	·048,576,
p_{41}= −	108·701,559,	$\bar{\theta}$= −	·470,126.

Further, from Σ_M, η=·303,024.

In deducing the product-moments *after they had been referred to the means*, the

* These products were in this case (as in all other cases) verified by calculating from the means of the arrays y_{x_p}, the expressions

$$S\left\{\frac{n_{x_p}y_{x_p}(x_p-\bar{x})}{N}\right\},\quad S\left\{\frac{n_x\,y_{x_p}(x_p-\bar{x})^2}{N}\right\},\quad S\left\{\frac{n_{x_p}y_{x_p}(x_p-\bar{x})^3}{N}\right\},\quad S\left\{\frac{n_{x_p}y_{x_p}(x_p-\bar{x})^4}{N}\right\}.$$

Of course it is easiest to calculate these products about some arbitrary origin coinciding with the abscissa of one array. If these products be then p'_{11}, p'_{21}, p'_{31}, p'_{41}, and $\bar{x}'$ be the mean, we have

$$p_{11}=p'_{11},$$
$$p_{21}=p'_{21}-2\bar{x}'p'_{11},$$
$$p_{31}=p'_{31}-3\bar{x}'p'_{21}+3\bar{x}'^2p'_{11},$$
$$p_{41}=p'_{41}-4\bar{x}'p'_{31}+6\bar{x}'^2p'_{21}-4\bar{x}'^3p'_{11},\ \ldots.$$

proper SHEPPARD's corrections were introduced. These are, if $\{p_{11}\}$, $\{p_{21}\}$, $\{p_{31}\}$, $\{p_{41}\}$ represent the uncorrected moments :—

$$p_{11}=\{p_{11}\}, \qquad p_{21}=\{p_{21}\},$$
$$p_{31}=\{p_{31}\}-\tfrac{1}{4}\{p_{11}\}, \qquad p_{41}=\{p_{41}\}-\tfrac{1}{2}\{p_{21}\},$$

the units of grouping being the units throughout.

From the constants for the arrays, I found

$$\chi_1-1=-\cdot000{,}675, \qquad \chi_2=-\cdot007{,}198.$$

Whence the probable error of η was determined by (xxxiii.). Its value was*

$$\text{Probable error of } \eta=\cdot012{,}913.$$

If found from the simple formula $\cdot67449\,(1-\eta^2)/N$, the value is $\cdot012{,}851$. We accordingly are again forced to the conclusion that η may for practical purposes be found from this simple formula, instead of the complicated result (xxxiii.). Although both χ_1-1 and χ_2 are small, it is very doubtful whether we can legitimately consider the system as homoscedastic. The dotted line ab of Diagram II. would fairly well represent increasing variability with age. The skewness of the arrays is relatively small and changes sign so frequently, that we can certainly not attribute any law to such heteroclitic tendencies as there are. They are probably due to errors of random sampling from truly isocurtic material.

It will be seen that the height frequencies with $\beta'_1=\cdot0160$ and $\beta'_2=3\cdot0815$ do not differ very much from a normal distribution ; in fact, we can lay no stress on the heteroclisy of the system at all. But the values of the standard deviations of the arrays, or the graph of σ_{n_x}/σ_y, certainly shows increasing variation with increasing age, a phenomenon with which one is familiar in a variety of other human characters.†

This heteroscedasticity, due to increasing variation with growth, would hardly have been anticipated from a mere inspection of the smallness of χ_1 ; it is somewhat obscured by the irregular values of the standard deviations of the small arrays at the adult end of the age range. The mean value of the standard deviation of the weighted arrays is $\sigma_y\sqrt{1-\eta^2}=3\cdot2992$ in 2-millim. units.

We now turn to the regression curves to see how far the conditions for the different types are satisfied. We have

$$\eta^2-r^2=\cdot005{,}312,$$
$$\phi_2\,(\eta^2-r^2)-\bar{\epsilon}^2=\cdot004{,}030,$$
$$\phi_2\,(\eta^2-r^2)-\bar{\epsilon}^2-(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2/(\phi_2\phi_4-\phi_3^2)=\cdot000{,}604.$$

* The contributions of the successive terms of (xxxiii.) are in fact given by

$$\Sigma_\eta^2=\frac{1}{N}\{\cdot824{,}785+\cdot001{,}870+\cdot004{,}673-\cdot000{,}472+\cdot001{,}888\}.$$

† See PEARSON : 'The Chances of Death and other Studies of Evolution,' vol. I., pp. 296, 307, 310, 314.

But the first should be zero, if the regression be linear; the second, if it be parabolic; and the third, if it be cubical.

We see increasing approximation to fulfilment of the several conditions. Referred to axes through the mean age and head height, the following are the regression curves* :—

(*a.*) *Straight line:*

$$Y_{x_p} = \cdot 662{,}979\, X_p.$$

(*b.*) *Parabola* (from equation (lxv.)):

$$Y_{x_p} = \cdot 055{,}749 + \cdot 667{,}570\, X_p - \cdot 041{,}001\, X_p^2.$$

(*c.*) *Cubic* (from equation (lvi.)):

$$Y_{x_p} = \cdot 280{,}194 + \cdot 722{,}886\, X_p - \cdot 029{,}580\, X_p^2 - \cdot 002{,}223\, X_p^3.$$

(*c'.*) *Cubic* (from equation (lix.)):

$$Y_{x_p} = \cdot 296{,}076 + \cdot 812{,}249\, X_p - \cdot 028{,}004\, X_p^2 - \cdot 005{,}740\, X_p^3.$$

(*c'*) will not give as good results as (*c*), for it depends on a use of the condition (lvii.) which is not absolutely fulfilled.

The following table gives the values in the case of the four curves :—

TABLE IV.—y_{x_p} = Mean Auricular Height of Girl's Head at Given Age.

x_p = age.	Regression line.	Regression parabola.†	Cubic (*c*).	Cubic (*c'*).	Observed.
3·5	117·95	114·49	116·90	118·94	115·25
4·5	118·61	115·87	117·66	118·94	116·96
5·5	119·27	117·17	118·42	119·16	117·47
6·5	119·94	118·39	119·24	119·57	119·10
7·5	120·60	119·52	120·08	120·14	120·30
8·5	121·26	120·57	120·93	120·84	121·63
9·5	121·92	121·55	121·78	121·62	121·72
10·5	122·59	122·43	122·62	122·45	122·82
11·5	123·25	123·24	123·42	123·26	123·14
12·5	123·91	123·97	124·18	124·15	123·89
13·5	124·58	124·61	124·88	124·95	124·86
14·5	125·24	125·17	125·52	125·65	125·71
15·5	125·90	125·65	126·07	126·22	126·16
16·5	126·57	126·05	126·52	126·68	126·53
17·5	127·23	126·36	126·87	126·93	126·91
18·5	127·89	126·59	127·09	126·96	127·02
19·5	128·55	126·75	127·18	126·74	129·56
20·5	129·22	126·81	127·11	126·22	123·82
21·5	129·88	126·80	126·88	125·38	126·50
22·5	130·54	126·71	126·48	124·28	125·25

* Y_{x_p} is here measured in millimetres and X_p in years.

† The maximum ordinate is at vertex of parabola, *i.e.*, $x = 8{\cdot}1409$, or age 20·84; its magnitude = 126·82.

An examination of this table and the graphs on Diagram II. seem to show:—

(i.) That cubic (c) is considerably better than cubic (c').

(ii.) That we do get a sensible betterment in passing from parabola to cubic, and, accordingly, that we must use in this the cubic to effectively describe the regression within the range of observation. Probably neither cubic nor parabola would effectively serve for extrapolation even close to the limits of observation.

Thus the cubic (c') starting at 3–4 with its point of inflection is clearly inadmissible, and the drop after 20 or 21 years of age, shown by both parabola and cubic, is, of course, only due to the anomalous character of the few girls over 18 left in the schools. Actually the shrinkage of measurements does not begin till at least 26 years, and is then far more gradual than these curves indicate.

But, as in all fitting of this kind, we obtain the best fit we can within the range, entirely at the expense of what may occur just outside the range. For this reason, as E. Perrin* has pointed out, a good interpolation curve is usually a bad extrapolation curve.

We might sum up our results for auricular height with age in girls by saying: That the correlation is non-linear, effectively cubic; heteroscedastic, there being increasing variability with growth; that while the total height frequency is not very far from normal the array frequencies are slightly heteroclitic, but so very irregular in sign, that probably we are dealing with a case of isocurtic homoclisy, to which the sparsity of data in the extreme arrays gives an appearance of anomic heteroclisy.

(10.) *Illustration* C.—*On the Skew Correlation between Size of Cell and Size of Body in* Daphnia magna.

Dr. E. Warren has dealt with this point in a memoir published in 'Biometrika,' vol. II., pp. 255–9. The resulting regression curve of size of cell for given size of body is very far from linear, and it is quite clear that the correlation is skew. It has already been noted in 'Biometrika' that the relationship is considerably obscured by the irregularities produced by ecdysis. Our object at present, however, is purely theoretical, namely, to show how a certain system of constants and of curves describes the actual correlationship, and for this purpose Dr. Warren's observations form as good material for graduation as we could expect to find. The following Table V. gives the observations with the working scales attached. I must refer to Dr. Warren's paper (p. 256) for the relation between the units of grouping on the working scales and those of the actual measurements on body and cell lengths. As far as correcting the raw moments is concerned, Sheppard's corrections were used for the cell sizes, but not for the body lengths, because the number of individuals in the latter case was perfectly arbitrary and there is no approach to high contact. The

* 'Biometrika,' vol. III., p. 99.

product moments were also uncorrected. The product moments were found in both ways (see p. 35, footnote) and the results thus verified.

Table V. gives the means, standard deviations, and third moments of the arrays; the latter are all small and superficially irregular in sign. I think we may say that there is no marked and continuous heteroclisy. On the other hand, I think we may say that while the clitic curve deviates to and fro from a zero base, the scedastic curve would fit better to a parabolic curve than to the straight line which is its mean. In other words, the variability of the cells increases with size of body (*i.e.*, growth) up to a certain stage and then decreases again. This result is obscured by the fall of the variability after each ecdysis. Roughly the ecdyses produce a rhythm in all three curves, the regression curve, the scedastic curve, and the clitic curve. When the means of the arrays are above the regression cubic, then the ordinates of the scedastic curve are above their mean and those of the clitic curve show positive skewness; when they are below the regression curve, we have lessened variability and negative skewness. In other words, the ecdyses are accompanied by lessened cell variability and negative skewness of distribution. I think we may state that there is a nomic heteroscedasticity due to growth of body, giving first an increased variability with growth and afterwards a decrease with age. There is probably isocurtic homoclisy. Both of these are, however, obscured by a semi-rhythmic heteroscedasticity and heteroclisy introduced by the ecdyses.

We now turn to the constants of the cell and body length distributions, merely noting that all these constants are given in terms of the units of the working scales.

Cell Constants.		*Body Length Constants.*	
Mean cell =	9·268,657,	Mean body length =	8·502,488,
σ_y =	2·541,734,	σ_x =	3·864,784,
μ_2 =	6·460,410,	ν_2 =	14·936,562,
μ_3 =	2·142,362,	ν_3 =	− 5·125,806,
μ_4 =	123·921,496,	ν_4 =	432·769,533,
		ν_5 =	− 425·276,682,
		ν_6 =	15192·5375,
β_1' =	·017,021,	β_1 =	·007,885,
β_2' =	2·969,111.	β_2 =	1·939,793,
Further		β_3 =	·043,796,
		β_4 =	4·559,091,
Σ_M =	1·454,600,	$\sqrt{\beta_1}$ =	− ·088,798,
λ_2 =	2·115,862,	ϕ_2 =	·931,908,
λ_4 =	15·142,840.	ϕ_3 =	− ·232,167,
Hence $(\lambda_4 - 3\lambda_2^2)/(4\lambda_2^2)$ =	·095,615.	ϕ_4 =	·788,409.

TABLE V.—Correlation between Body and Cell Lengths in *Daphnia magna*.

Body length.	Size of cell. 1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	Totals.	Means.	Standard deviation.	m_3.
1	—	—	1	2	9	6	2	—	—	—	—	—	—	—	—	—	—	20	5·3000	0·9092	−0·2460
2	1	2	8	19	37	34	23	15	10	1	—	—	—	—	—	—	—	150	5·8333	1·6620	−0·8209
3	—	—	1	2	5	13	23	25	13	11	7	—	—	—	—	—	—	100	7·7900	1·6860	−0·3974
4	—	—	—	3	7	13	13	11	6	15	8	4	—	—	—	—	—	80	8·0500	2·1539	+0·5415
5	—	—	2	2	5	9	11	29	28	16	13	13	9	7	3	3	—	150	9·4733	2·6694	+4·5706
6	—	—	—	2	6	9	23	33	29	17	20	—	1	—	—	—	—	140	8·4357	1·7087	−0·8367
7	—	—	—	3	12	19	34	43	41	37	26	11	4	—	—	—	—	230	8·5957	1·9384	−0·3748
8	—	—	—	—	—	3	8	11	18	28	20	16	11	2	3	—	—	120	10·2667	1·9439	+0·5079
9	—	—	—	1	3	6	8	12	24	26	29	26	22	14	3	6	—	180	10·7611	2·4469	−2·7079
10	—	—	—	—	1	3	11	17	16	17	21	19	12	17	8	5	3	150	11·0267	2·6949	+2·4790
11	—	—	—	—	—	1	8	13	27	25	20	14	17	10	8	4	3	150	10·9533	2·4470	+6·5234
12	—	—	—	2	4	2	17	25	31	30	16	9	4	—	—	—	—	140	9·1000	1·7947	−1·4494
13	—	—	—	—	6	8	11	19	27	21	17	10	1	—	—	—	—	120	9·0000	1·8619	−1·8000
14	—	—	—	—	1	4	22	19	34	49	45	27	11	5	3	—	—	220	10·0364	1·8894	+0·0196
15	—	—	—	—	—	1	2	6	6	15	15	13	2	—	—	—	—	60	10·3167	1·5165	−2·1790
Totals . .	1	2	12	36	96	131	216	278	310	308	257	162	94	55	28	18	6	2010	9·2687	2·5417	+2·1424

We have next the product moments referred to the means

$$\begin{aligned} p_{11} &= \quad 3{\cdot}892{,}863, & \text{whence} \quad r &= \quad {\cdot}394{,}862, \\ p_{21} &= -\ 12{\cdot}104{,}322, & \bar{\epsilon} &= -{\cdot}281{,}831, \\ p_{31} &= \quad 127{\cdot}348{,}064, & \zeta &= \quad {\cdot}098{,}578, \\ p_{41} &= -541{\cdot}433{,}455, & \bar{\theta} &= -{\cdot}759{,}344. \end{aligned}$$

Further, from Σ_M,

$$\eta = {\cdot}572{,}287.$$

From the constants for the arrays I deduced

$$\chi_1 - 1 = -{\cdot}108{,}148, \qquad \chi_2 = {\cdot}088{,}323.$$

These are higher values of $\chi_1 - 1$ and χ_2 than we have found in the first two illustrations.

We now obtain, showing the contribution of each term of (xxxiii.),

$$\Sigma_\eta^2 = \frac{1}{N}\{{\cdot}452{,}240 - {\cdot}002{,}528 + {\cdot}010{,}803 - {\cdot}013{,}180 - {\cdot}027{,}875\}.$$

Whence probable error of $\eta = {\cdot}67449\,\Sigma_\eta = {\cdot}0097$.

Had we calculated the probable error of η from (xxxiv.), we should have found it equal to ·0101. The difference is greater than in the two previous illustrations, but is only ·0004, and this would have no significance in any practical use of the probable error. We again conclude, therefore, that (xxxiv.) is sufficiently close to replace (xxxiii.) in practice.

For the mean standard deviation of the weighted arrays we have

$$\sigma_{a_y} = \sigma_y \sqrt{1-\eta^2} = 2{\cdot}084{,}358.$$

If we now examine the criteria for the nature of the regression, we have

$$\begin{aligned} \eta^2 - r^2 &= {\cdot}171{,}596, \\ \phi_2(\eta^2 - r^2) - \bar{\epsilon}^2 &= {\cdot}080{,}483, \\ \phi_2(\eta^2 - r^2) - \bar{\epsilon}^2 - (\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)^2/(\phi_2\phi_4 - \phi_3^2) &= {\cdot}079{,}457. \end{aligned}$$

We should conclude, therefore, that linear regression is inadmissible, but that parabolic or cubic will be moderately successful, the latter not very much better than the former. Our moderate success only in this case is, of course, due to the irregularity of the results to be graduated, the influence of the ecdyses being so disturbing that we really need a curve periodically varying from the graduated regression curve.

We have the following regression curves :—

(*a.*) *Straight line:*

$$Y_{x_p} = {\cdot}259{,}687\, X_p.$$

(*b.*) *Parabola* from (lxv.):

$$Y_{x_p}=1{\cdot}097{,}690+{\cdot}236{,}135\,X_p-{\cdot}073{,}490\,X_p^2.$$

The maximum occurs when $X_p=1{\cdot}6066$, and is given by $Y_{x_p}=1{\cdot}2874$, thus occurring within the limits of observation.*

(*c.*) *Cubic* from (lix.):

$$Y_{x_p}={\cdot}752{,}856+{\cdot}193{,}058\,X_p-{\cdot}049{,}817\,X_p^2+{\cdot}001{,}710\,X_p^3.$$

In all these cases Y_{x_p} and X_p are measured from the means of the cell and body lengths, or from 9·268,657 and 8·502,488 respectively.

Table VI. gives the calculated and observed results, and the whole system is represented in Diagram III. Either the parabola or cubic graduates quite well the results, allowing for the periodic deviation, and we may fairly describe the system as a heteroscedastic cubic regression with isocurtic homoclisy. The correlation ratio is very sensibly different from the correlation coefficient. The regression cubic does not differ widely from that given in ' Biometrika,' which was obtained without weighting the means of the arrays, and by simply striking the best cubic of the given type through the points.

TABLE VI.—y_{x_p}=Mean Cell Length for Given Body Length in *Daphnia*.

x_p = body length.	Regression line.	Regression parabola.	Regression cubic.	Observed.
1	7·320	4·458	5·047	5·300
2	7·580	5·724	6·190	5·833
3	7·840	6·842	7·166	7·790
4	8·099	7·813	7·986	8·050
5	8·359	8·638	8·661	9·473
6	8·619	9·315	9·200	8·436
7	8·879	9·846	9·613	8·596
8	9·138	10·229	9·912	10·267
9	9·398	10·466	10·105	10·761
10	9·658	10·555	10·205	11·027
11	9·917	10·498	10·220	10·953
12	10·177	10·293	10·161	9·100
13	10·437	9·942	10·038	9·000
14	10·696	9·443	9·861	10·036
15	10·956	8·798	9·642	10·317

(11.) *Illustration* D.—*On the Skew Correlation between Number of Branches to the Whorl and Position of the Whorl on the Stem in* Equisetum arvense.

I have selected this example not on account of any biological importance, because the material is—especially with regard to the first and last two whorls—unsatisfactory either on account of irregularity or of insufficiency of material. It has been taken

* Actual values on working scales, $x_p=10{\cdot}1091$ and $y_{x_p}=10{\cdot}5560$.

purely from its statistical interest, because it gives a series with markedly skew correlation, having a regression curve of a rough **S**-shaped character. If we omit the first and last whorls, we get, as I have already shown,* a remarkably close fit with a cubical regression curve. My present object, however, is not to consider any law of growth, but merely a mass of statistical material, to be dealt with by the processes of the present paper.

We may anticipate that the irregularities of the series, indicated in the memoir just referred to, will make themselves manifest in a less satisfactory fitting of the regression curve than occurs when we deal with the more homogeneous group of equally weighted whorls fitted in the diagram of that paper. Table VII. gives the data, with the means, standard deviations, and third moments of each array.

The axis of x shall be taken to give the position of the whorl on the stem and that of y to denote the number of branches. We require the regression curve of y on x, or the probable number of branches on a whorl in a given position. We shall not use SHEPPARD's corrections for the moments of either the x or y-characters, as high contact certainly does not hold for both at the low-value ends of their ranges.

We have the following constants:—

Position Constants.		*Branch Constants.*	
Mean position =	6·403,315,	Mean number of branches =	7·216,851,
σ_x =	3·542,604,	σ_y =	3·278,499,
ν_2 =	12·550,046,	μ_2 =	10·748,557,
ν_3 =	8·249,534,	μ_3 =	− 24·313,478,
ν_4 =	319·515,824,	μ_4 =	245·811,660,
ν_5 =	644·095,176,		
ν_6 =	11203·5814,		
β_1 =	·034,429,	β'_1 =	·476,044,
β_2 =	2·028,625,	β'_2 =	2·127,658.
β_3 =	·214,190,	Further	
β_4 =	5·667,884,	Σ_M =	2·789,949,
$\sqrt{\beta_1}$ =	·185,550,	λ_2 =	7·783,815,
ϕ_2 =	·994,196,	λ_4 =	140·441,685.
ϕ_3 =	·592,384,	Hence	
ϕ_4 =	1·518,136.	$(\lambda_4 - 3\lambda_2^2)/(4\lambda_2^2)$ =	−·170,503.

We have next the product moments referred to the means

* 'Proc. Roy. Soc.,' vol. 71, p. 308.

Table VII.

Position of Whorl.	Number of branches to the whorl. 1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	Totals.	Mean.	Standard deviation.	Third moment.
1	2	2	3	10	9	8	13	29	22	17	11	—	—	126	7·619	2·360	− 9·437
2	—	—	—	—	1	3	5	21	37	40	16	3	—	126	9·294	1·273	− 1·111
3	—	—	—	—	—	—	9	9	35	45	23	5	—	126	9·627	1·187	− 0·650
4	—	—	—	—	—	—	6	10	33	45	28	3	1	126	9·730	1·151	− 0·464
5	—	—	—	—	—	—	8	10	35	41	30	2	—	126	9·643	1·158	− 0·780
6	—	—	—	—	2	3	6	13	35	38	24	3	—	124	9·427	1·375	− 2·171
7	—	—	1	4	2	6	12	23	28	29	17	1	—	123	8·732	1·781	− 5·013
8	2	3	7	5	5	13	21	23	24	14	4	—	—	121	7·297	2·291	− 9·727
9	8	10	13	14	9	14	19	17	10	5	—	—	—	119	5·555	2·553	− 2·693
10	18	20	13	11	17	14	11	5	1	—	—	—	—	110	3·964	2·199	+ 2·455
11	31	29	18	9	5	3	1	1	—	—	—	—	—	97	2·443	1·506	+ 4·392
12	24	34	6	2	—	—	1	—	—	—	—	—	—	67	1·866	0·960	+ 2·210
13	24	14	—	—	1	—	—	—	—	—	—	—	—	39	1·462	0·746	+ 1·132
14	8	4	—	—	—	—	—	—	—	—	—	—	—	12	1·333	0·471	+ 0·707
15	3	1	—	—	—	—	—	—	—	—	—	—	—	4	1·250	0·433	+ 0·094
16	2	—	—	—	—	—	—	—	—	—	—	—	—	2	1·000	0·000	+ 0·000
Totals . .	122	117	61	55	51	64	112	161	260	274	153	17	1	1448	7·217	3·278	−24·313

$$p_{11} = -\ 8{\cdot}225{,}585, \qquad \text{whence} \quad r = -{\cdot}708{,}222,$$
$$p_{21} = -\ 21{\cdot}471{,}321, \qquad \bar{\epsilon} = -{\cdot}390{,}436,$$
$$p_{31} = -205{\cdot}084{,}042, \qquad \bar{\zeta} = +{\cdot}029{,}733,$$
$$p_{41} = -917{\cdot}984{,}938, \qquad \bar{\theta} = -{\cdot}960{,}212.$$

Further, from Σ_M,

$$\eta = {\cdot}850{,}984.$$

From the constants for the arrays we deduce

$$\chi_1 - 1 = -{\cdot}356{,}367, \qquad \chi_2 = -{\cdot}312{,}952.$$

We now obtain, showing the contribution of each term of (xxxiii.),

$$\Sigma_\eta^{\,2} = \frac{1}{N}\{{\cdot}076{,}080 - {\cdot}157{,}932 + {\cdot}055{,}359 + {\cdot}079{,}662 + {\cdot}038{,}579\}.$$

Whence probable error of $\eta = {\cdot}67449\ \Sigma_\eta = {\cdot}0054$.

Had we calculated the probable error ot η from (xxxiv.) we should have found it equal to ·0049. The difference ·0005 is not of importance for practical purposes. Yet in this case it is clear that the values of $\chi_1 - 1$ and χ_2 are very sensible. Thus we see that a very marked heteroscedastic and heteroclitic system with continuously changing standard deviation and skewness scarcely affects for practical purposes (*i.e.*, to three significant figures) the probable error of η. All four of our illustrations therefore confirm the conclusion that:

For practical purposes the probable error of the correlation ratio, η, may be taken as ·67449 $(1-\eta^2)/N$.

Our Diagram IV. gives the values of the relative standard deviations of the arrays, or, σ_{n_x}/σ_y, the horizontal line giving $\sqrt{1-\eta^2} = {\cdot}5252$, or the mean value of the relative standard deviations of the weighted arrays. We have also the clitic curve giving $\frac{1}{2}\sqrt{\beta_1}$, for each array.* The remarkable smoothness of these scedastic and clitic curves in this case indicates how far certain types of correlation surfaces diverge from pure normality of distribution, the divergence being obviously nomic.

We now turn to the regression curves and write down the conditions for the different types; the three expressions should be zero for linear, parabolic, and cubical regression respectively

$$\eta^2 - r^2 = {\cdot}222{,}596,$$
$$\phi_2(\eta^2 - r^2) - \bar{\epsilon}^2 = {\cdot}068{,}864,$$
$$\phi_2(\eta^2 - r^2) - \bar{\epsilon}^2 - (\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)^2/(\phi_2\phi_4 - \phi_3^2) = {\cdot}010{,}127.$$

* $\frac{1}{2}\sqrt{\beta_1}$ = difference between mode and mean divided by standard deviation = skewness in the case of skew-curves of Type III. ('Phil. Trans.,' A, vol. 186, p. 373), and may be taken as a reasonable measure of the skewness for those cases in which the fuller form involving β_2 would involve too laborious calculations. If in equation (xii.) of the present memoir we put $\beta_2 = 3 +$ a small quantity, and remember that β_1 is itself a small quantity, we see that the more correct formula for the skewness involving β_2 reduces, neglecting terms of 2nd order, to $\frac{1}{2}\sqrt{\beta_1}$.

We see at once that the straight line is inadmissible, the parabola will not be very good, and the cubic only moderately appropriate. The conditions are not nearly so closely fulfilled as in the cases of woodruff and head heights; the last two are better than in the case of Daphnia cells, but while the deviations in the case of Daphnia were irregular, there being no approximate smoothness in the scedastic or clitic curves, we shall find here more uniform deviations which would probably be partially allowed for by a quartic regression curve.

The following are the regression curves:—

(*a.*) *Straight line:*

$$Y_{x_p} = -\cdot 655{,}423\, X_p.$$

(*b.*) *Parabola* from (lxv.):

$$Y_{x_p} = 1\cdot 551{,}307 - \cdot 574{,}171\, X_p - \cdot 123{,}610\, X_p^2.$$

The maximum ordinate is at the position $X_p = -2\cdot 3225$, or $x_p = 4\cdot 0808$, with maximum number of branches $y_p = 9\cdot 435$.

(*c.*) *Cubic* from (lvi.):

$$Y_{x_p} = 1\cdot 590{,}413 - \cdot 987{,}694\, X_p - \cdot 137{,}641\, X_p^2 + \cdot 016{,}605\, X_p^3.$$

In all cases X_p and Y_{x_p} are measured from the mean position and the mean number of branches, *i.e.*, 6·403,315 and 7·216,851 respectively.

The following table contains the calculated and observed results:—

TABLE VIII.—Mean Number of Branches to each Whorl in *Equisetum.*

Position.	Regression line.	Regression parabola.	Regression cubic.	Observed.	Regression cubic without first whorl.
1	10·758	8·262	7·506	7·619	[8·207]
2	10·103	8·900	9·070	9·294	8·929
3	9·447	9·291	9·920	9·627	9·869
4	8·792	9·434	10·156	9·730	10·161
5	8·137	9·330	9·876	9·643	9·911
6	7·481	8·980	9·182	9·427	9·224
7	6·826	8·382	8·172	8·732	8·205
8	6·170	7·536	6·947	7·297	6·962
9	5·515	6·444	5·605	5·555	5·599
10	4·859	5·104	4·247	3·964	4·223
11	4·204	3·517	2·971	2·443	2·939
12	3·549	1·683	1·879	1·866	1·854
13	2·893	−0·399	1·069	1·462	1·072
14	2·238	−2·727	0·641	1·333	0·700
15	1·582	−5·303	0·694	1·250	0·844
16	0·927	−8·126	1·328	1·000	1·610

In the last column I have placed the results of re-working the whole system, omitting the first whorl as largely influenced by the ground condition at the foot of

the stem.* The improvement of fit is not sufficiently great to justify a publication of all the constants for the distribution in this modified case. But there is improvement for the higher whorls, which are so few in number as to be wholly insignificant when compared with the weight of the first few low whorls.

It will be noticed at once that the line and the parabola (which gives at the top of the stem negative numbers!) are absolutely unsuitable for representing the facts of the case. The cubic is better and certainly gives the general trend of the observations, but in this our last illustration we have clearly reached the limit of material to which such cubical regression can be satisfactorily applied. See Diagram V.

(12.) *Quartic Regression.*

It seemed of some interest in this case of *Equisetum* to ascertain whether any real improvement in description would be reached by considering the quartic regression curve. I briefly indicate the theory in this case as developed from the general method in the footnote, p. 25. We shall now have

$$Y_{x_p}/\sigma_y = b_0 + b_1(X_p/\sigma_x) + b_2(X_p/\sigma_x)^2 + b_3(X_p/\sigma_x)^3 + b_4(X_p/\sigma_x)^4.$$

Eliminating b_0 and b_1, by the processes familiar to us from the case of cubical regression, we have

$$\begin{aligned} Y_{x_p}/\sigma_y = {} & r(X_p/\sigma_x) + b_2\{(X_p/\sigma_x)^2 - \sqrt{\beta_1}(X_p/\sigma_x) - 1\} \\ & + b_3\{(X_p/\sigma_x)^3 - \beta_2(X_p/\sigma_x) - \sqrt{\beta_1}\} \\ & + b_4\{(X_p/\sigma_x)^4 - (\beta_3/\sqrt{\beta_1})(X_p/\sigma_x) - \beta_2\} \quad \text{(lxx.)}. \end{aligned}$$

Hence as before

$$\left.\begin{aligned} \bar{\epsilon} &= b_2\phi_2 + b_3\phi_3 + b_4\phi_5 \\ \bar{\zeta} &= b_2\phi_3 + b_3\phi_4 + b_4\phi_6 \\ \bar{\theta} &= b_2\phi_5 + b_3\phi_6 + b_4\phi_7 \end{aligned}\right\} \quad \text{(lxxi.)},$$

where ϕ_2, ϕ_3, and ϕ_4 are given as before by (li. and liv.), while

$$\phi_5 = \beta_4 - \beta_3 - \beta_2 \quad \text{(lxxii.)},$$

$$\phi_6 = (\beta_5 - \beta_2\beta_3 - \beta_2\beta_1)/\sqrt{\beta_1} \quad \text{(lxxiii.)},$$

$$\phi_7 = (\beta_1\beta_6 - \beta_3^2 - \beta_1\beta_2^2)/\beta_1 \quad \text{(lxxiv.)},$$

and

$$\beta_5 = \nu_7\nu_3/\sigma_x^{10}, \quad \beta_6 = \nu_8/\sigma_x^8 \quad \text{(lxxv.)}.$$

Solving, we have

$$b_4 = \frac{\bar{\theta}(\phi_2\phi_4 - \phi_3^2) - \bar{\epsilon}(\phi_4\phi_5 - \phi_3\phi_6) - \bar{\zeta}(\phi_2\phi_6 - \phi_3\phi_5)}{\phi_2\phi_4\phi_7 - \phi_7\phi_3^2 - \phi_4\phi_5^2 - \phi_2\phi_6^2 + 2\phi_3\phi_5\phi_6} \quad \text{(lxxvi.)},$$

* 'Roy. Soc. Proc.,' vol. 71, pp. 308–310.

and

$$\left.\begin{aligned} b_2 &= \frac{\bar{\epsilon}\phi_4-\bar{\zeta}\phi_3}{\phi_2\phi_4-\phi_3^2} - b_4\frac{\phi_4\phi_5-\phi_3\phi_6}{\phi_2\phi_4-\phi_3^2} \\ b_3 &= \frac{\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3}{\phi_2\phi_4-\phi_3^2} - b_4\frac{\phi_2\phi_6-\phi_3\phi_5}{\phi_2\phi_4-\phi_3^2} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxxvii).}$$

Substituting in (lxx.), the solution is completed. The advantage of this form is that we see clearly the modifications made in b_2 and b_3 as we pass from cubical to quartic regression. On the other hand, ϕ_6 and ϕ_7, as shown by (lxxv.), involve the 7th and 8th moments of the x-character. These are not only very laborious to calculate, but, as we have already shown, are as a rule very untrustworthy.

If we proceed as on p. 26, equation (lvii.), we find

$$\eta^2-r^2=b_2\bar{\epsilon}+b_3\bar{\zeta}+b_4\bar{\theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(lxxviii.).}$$

Using this and not the third equation of (lxxi.), we replace (lxxvi.) by

$$b_4=(\phi_2\phi_4-\phi_3^2)\frac{\left\{\eta^2-r^2-\dfrac{\bar{\epsilon}^2}{\phi_2}-\dfrac{(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2}{\phi_2(\phi_2\phi_4-\phi_3^2)}\right\}}{\bar{\theta}(\phi_2\phi_4-\phi_3^2)-\bar{\epsilon}(\phi_4\phi_5-\phi_3\phi_6)-\bar{\zeta}(\phi_2\phi_6-\phi_3\phi_5)} \quad . \quad \text{(lxxix.).}$$

This equation for b_4 only involves the 7th and not the 8th moment, but like the corresponding form (lx.) suffers from being a ratio of small quantities. (lxxvii.) completes the solution as before.

(lxxvii.) and (lxxix.) in conjunction give us a necessary condition for quartic regression. We can indeed now write the whole series of conditions as follows:—

Linear regression:

$$\eta^2-r^2=0.$$

Parabolic regression:

$$\eta^2-r^2-\bar{\epsilon}^2/\phi_2=0.$$

Cubical regression:

$$\eta^2-r^2-\bar{\epsilon}^2/\phi_2-(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2/\{\phi_2(\phi_2\phi_4-\phi_3^2)\}=0.$$

Quartic regression:

$$\eta^2-r^2-\bar{\epsilon}^2/\phi_2-\frac{(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2}{\phi_2(\phi_2\phi_4-\phi_3^2)}-\frac{\{\bar{\theta}(\phi_2\phi_4-\phi_3^2)-\bar{\epsilon}(\phi_4\phi_5-\phi_3\phi_6)-\bar{\zeta}(\phi_2\phi_6-\phi_3\phi_5)\}^2}{(\phi_2\phi_4-\phi_3^2)(\phi_2\phi_4\phi_7-\phi_7\phi_3^2-\phi_4\phi_5^2-\phi_2\phi_6^2+2\phi_3\phi_5\phi_6)}=0$$

. . . . . (lxxx.).

We now have a third possibility: we can get rid of the fourth product moment $\bar{\theta}$ from the value of b_4 and write it:

$$b_4=\pm\sqrt{\frac{\eta^2-r^2-\bar{\epsilon}^2/\phi_2-(\bar{\zeta}\phi_2-\bar{\epsilon}\phi_3)^2/\{\phi_2(\phi_2\phi_4-\phi_3^2)\}}{\phi_7-\phi_5\dfrac{\phi_4\phi_5-\phi_3\phi_6}{\phi_2\phi_4-\phi_3^2}-\phi_6\dfrac{\phi_2\phi_6-\phi_3\phi_5}{\phi_2\phi_4-\phi_3^2}}} \quad . \quad . \quad \text{(lxxxi.).}$$

While this value of b_4 does not suffer like (lxxix.) from being the ratio of small quantities, and would *a priori* appear to save the calculation of $\bar{\theta}$, yet the right sign of the root may not be ovious on inspection, so that an actual determination of $\bar{\theta}$ to find the sign of b_4 may after all be needful. If (lxxx.) were absolutely satisfied, (lxxxi.), (lxxix.) and (lxxvi.) would lead to identical results; but this will rarely be true in practice. In any of the three cases b_2 and b_3 will be given by (lxxviii.). On the whole, I consider that (lxxxi.) and (lxxvi.) will give the better results, and probably the former the best, but it will generally require as much arithmetic as the latter.

(13). *Illustration* E.—*Calculation of the Quartic Regression Curve in the Case of* Equisetum arvense.

The only new constants required are:

$$\nu_7 = 43{,}207{\cdot}386, \qquad \text{whence } \beta_5 = 1{\cdot}144{,}882,$$

$$\nu_8 = 507{,}649{\cdot}540, \qquad \beta_6 = 20{\cdot}463{,}633,$$

and:

$$\phi_5 = 3{\cdot}425{,}069, \qquad \phi_6 = 3{\cdot}452{,}046,$$

$$\phi_7 = 15{\cdot}015{,}792.$$

These lead us to:

$$\frac{\phi_4\phi_5 - \phi_3\phi_6}{\phi_2\phi_4 - \phi_3^2} = 2{\cdot}723{,}384, \qquad \frac{\phi_2\phi_6 - \phi_3\phi_5}{\phi_2\phi_4 - \phi_3^2} = 1{\cdot}211{,}194,$$

$$\Delta_\phi = \begin{vmatrix} \phi_2, & \phi_3, & \phi_5 \\ \phi_3, & \phi_4, & \phi_6 \\ \phi_5, & \phi_6, & \phi_7 \end{vmatrix} = 1{\cdot}745{,}622.$$

Our successive conditions are therefore:

$$\eta^2 - r^2 = {\cdot}222{,}596,$$

$$\eta^2 - r^2 - \bar{\epsilon}^2/\phi_2 = {\cdot}069{,}266,$$

$$\eta^2 - r^2 - \bar{\epsilon}^2/\phi_2 - (\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)^2/\{\phi_2(\phi_2\phi_4 - \phi_3^2)\} = {\cdot}010{,}186,$$

$$\eta^2 - r^2 - \bar{\epsilon}^2/\phi_2 - (\bar{\zeta}\phi_2 - \bar{\epsilon}\phi_3)^2/\{\phi_2(\phi_2\phi_4 - \phi_3^2)\} - \frac{\{\bar{\theta}(\phi_2\phi_4 - \phi_3^2) - \bar{\epsilon}(\phi_4\phi_5 - \phi_3\phi_6) - \bar{\zeta}(\phi_2\phi_6 - \phi_3\phi_5)\}^2}{(\phi_2\phi_4 - \phi_3^2)\Delta_\phi} = {\cdot}007{,}200,$$

whence we see the successive approximations to the fulfilment of the conditions. Clearly great gains arise when we pass from linear to parabolic, and from parabolic to cubic regression, but the advance is not so conspicuous when we pass to quartic regression.

We have :—

From (lxxvi.): $b_4 = \cdot 044,517$, and $b_2 = -\cdot 648,122$, $b_3 = \cdot 171,260$,

From (lxxix.): $b_4 = \cdot 151,842$, and $b_2 = -\cdot 940,410$, $b_3 = \cdot 041,981$,

From (lxxxi.): $b_4 = \cdot 025,999$, and $b_2 = -\cdot 597,691$, $b_3 = \cdot 193,688$.

The equations to the three corresponding quartics are :

(*a*). $Y_{x_p} = 1\cdot 724,611 - \cdot 913.208\ X_p - \cdot 169,311\ X_p^2 + \cdot 012,629\ X_p^3 + \cdot 000,927\ X_p^4$,

(*b*). $Y_{x_p} = 2\cdot 047,717 - \cdot 734,966\ X_p - \cdot 245,667\ X_p^2 + \cdot 003,096\ X_p^3 + \cdot 003,161\ X_p^4$,

(*c*). $Y_{x_p} = 1\cdot 668,788 - \cdot 944,192\ X_p - \cdot 156,137\ X_p^2 + \cdot 014,283\ X_p^3 + \cdot 000,541\ X_p^4$.

The values of Y_{x_p} and X_p are as before measured from the means, or 7·216,851 and 6·403,315 respectively.

The values of the observed and calculated ordinates are given in Table IX., and the graph of the results in the lower half of Diagram V.

TABLE IX.—Mean Number of Branches to Whorl in *Equisetum* deduced from Quartic Regression.

Position.	Quartic (*a*).	Quartic (*b*).	Quartic (*c*).	Observed.
1	7·731	8·269	7·637	7.619
2	8·950	8·662	9·000	9·294
3	9·715	9·222	9·800	9·627
4	10·014	9·674	10·073	9·730
5	9·858	9·816	9·866	9·643
6	9·281	9·521	9·240	9·427
7	8·339	8·740	8·270	8·732
8	7·109	7·498	7·042	7·297
9	5·692	5·898	5·656	5·555
10	4·209	4·116	4·225	3·964
11	2·816	2·407	2·875	2·443
12	1·651	1·100	1·745	1·866
13	0·930	0·600	0·987	1·462
14	0·857	1·389	0·766	1·333
15	1·665	4·022	1·259	1·250
16	3·609	9·133	2·657	1·000

From these results we deduce the following conclusions :—

(i.) That the use of a quartic instead of a cubic regression curve has not very markedly bettered the fit. The failure to get a closer fit lies largely in the nature of the material. The number of plants with more than 13 whorls is very few, and their contribution allows little weight to the tail of the regression curve. Further, all our

attempts to fit a smooth regression curve show that the observed data are unduly flattened at the top. If we confine ourselves to a homogeneous series of 110 plants with ten whorls apiece, we get a remarkably good fit.* The **S**-shape of the regression line as indicated in both cubic and quartic does, however, appear to be characteristic of the nature of the plant, and I take it that more ample material would allow of a closer analytical description by a simple cubic. I doubt whether for practical statistics the use of the quartic will often be requisite.

(ii.) The comparative failure of the quartic (*b*) shows us that a formula like (lxxix.) is of small service. This corresponds fully to our experience in the use of (lx.) in the case of the cubic. In both cases we get rid of a high moment by making a certain constant the ratio of two small quantities, and experience shows us that the result is unsatisfactory. It is accordingly preferable to use formulæ involving high moments of one variable in preference to those with a ratio of small quantities.

(iii.) The quartic (*c*) appears as good, if not slightly better, than quartic (*a*). In (*c*) we have got rid of a high product moment, $\bar{\theta}$, by supposing the quartic condition (lxxx.) rigidly fulfilled. This of course is not the case. It is clear that product moments like $\bar{\theta}$ of the 5th order are far from advantageous, and this is the same principle which was in evidence when we found (lxv.) giving better results than (lxiv.) for parabolic regression. Hence we must further conclude that the use of third, fourth or fifth product moments is disadvantageous as compared respectively with fifth to eighth moments of one variable. Or, a moment two degrees higher is preferable to a product moment in calculating correlation values. This is, I think, consonant with our knowledge of the relative magnitude of the probable errors in the two cases.

(14.) *General Conclusions.*

(i.) The present paper provides us with a general method of dealing with the regression line and the variability of arrays in the case of skew correlation, without any assumption as to the analytical form of the skew correlation surface.

(ii.) It provides a nomenclature and classification of the types of array variability which may be of service.

Arrays are either *homoclitic* or *heteroclitic*, according as their skewnesses are of equal magnitude or not. Arrays are further *homoscedastic* or *heteroscedastic*, according as their standard deviations are alike or different. Skew arrays are termed *allocurtic;* if arrays are symmetrical about their mean, they are *isocurtic.*

A heteroclitic system of arrays may be *nomic* or *anomic*, according as the skewness of the arrays changes continuously or irregularly with the position of the array.

A heteroscedastic system of arrays is also either *nomic* or *anomic*, according as the standard deviation of the arrays changes continuously or irregularly with the

* 'Roy. Soc. Proc.,' vol. 71, p. 308.

position of the arrays. Anomic heteroclisy and anomic heteroscedasticity probably only signify that our material is either heterogeneous or too sparse to free us from the large errors of random sampling in the extreme arrays. Still the terms will be found of use in describing the actual data.

The curve in which the skewness of the array is plotted to its position is termed the *clitic curve;* the curve in which the ratio of the standard deviation of the array to the standard deviation of the character in the population at large is plotted to position is termed a *scedastic curve.*

(iii.) The types of regression have been classified into *linear*, *parabolic*, *cubic* and *quartic*. For most practical purposes the first three suffice. Necessary criteria have been given for each case. But as in the case of the skew frequency of one character, an indefinite number of conditions ought theoretically to be fulfilled. Practically in dealing with frequency, no criteria are absolutely fulfilled, and the probable errors of the expressions used become unmanageable as we ascend in the scale. We must therefore be content to estimate the degree of approximation with which one or two necessary criteria are satisfied.

The fundamental test of deviation from the familiar form of linear regression is the inequality of the correlation coefficient r and the newly introduced correlation ratio η. The probable error of this latter is determined. It is shown that $\sigma_y \sqrt{1 - \eta^2}$ is the mean standard deviation of a system of arrays in skew correlation. The ease with which η can be calculated suggests that in many cases it should accompany, if not replace the determination of the correlation coefficient.

In the determination of the constants of the regression curve we must use moments and product moments. The limitations to the order of the curve used depend: (a) on the labour of the arithmetic, (b) on the increasing probable errors of the higher moments and product moments. For these reasons it seems idle to propose going beyond the 6^{th} to 8^{th} moments, or the 3^{rd} to 5^{th} product-moments. Practical experience suggests that little is to be gained by using moments beyond the 6^{th}, or product moments beyond the 3^{rd}. A quartic regression curve may be useful occasionally, but it has yet to justify its necessity. As our object is not to reproduce the given data, but to provide a graduation for them, which smooths down the errors of random sampling, we believe that any legitimate and practical theory must discard the high moments and high product moments with which THIELE and LIPPS propose to deal.

(iv.) There is one point to which reference ought to be made. Some reader may enquire why the method of my paper on curving fitting* should not be applied to these regression curves *in general*, as we have in practice once or twice already applied it. It would seem that that method is the easier, involving in the case of the quartic only quantities analogous to our r, e, ζ and θ. The answer is

* "On the Systematic Fittings of Curves to Observations a d Measurements." 'Biometrika,' vol. I., pp. 265–303, and vol. II., pp. 1–23, especially the latter, pp. 11–15.

straightforward: that process supposes every y_{x_p} to have equal weight, or n_{x_p} to be the same for each array. Hence the higher moments of the x-character, which are really involved, can be written down without calculation once and for all.* The complexity of our present investigation arises from the introduction of the weighting into the calculation of the moments of the x-character, as well as into that of the product moments r, e, ζ, θ. Our results therefore, although they might not look so good on a graph of the regression curve, would be markedly better, if due weight were given to the frequency of each array. The difference of the two conceptions is comparable to the determination of the regression on the one hand from the correlation coefficient, and on the other from merely striking a line through the plotted means of the arrays. The method of moments in the present case, if we except the use of η, is identical with that of fitting a curve to a *continuum* in space by the method of least squares.

(v.) No stress whatever is laid on the actual instances here selected for illustration of the methods of this paper. I have merely chosen out of available material cases in which I had come across skew regression of various types. Thus we find :—

(*a.*) The correlation of the number of branches and position of the whorl in *Asperula odorata* is practically parabolic, homoscedastic and of nomic heteroclisy.

(*b.*) The correlation between auricular height of head and age in girls is cubical, of nomic heteroscedasticity and of anomic heteroclisy. It is probably really a case of isocurtosis.

(*c.*) The correlation of size of cell and size of body in *Daphnia magna*, allowing for the irregularities produced by the ecdyses, is parabolic or cubic, of nomic heteroscedasticity, and probably, but for the above-mentioned irregularities, of isocurtic homoclisy.

(*d.*) The correlation of the number of branches and position of the whorl in *Equisetum arvense* is cubical or possibly even quartic, of markedly nomic heteroscedasticity and markedly nomic heteroclisy.

It is not impossible that slips have occurred in the lengthy arithmetic involved, but every important piece of work has been done independently twice, once by Dr. Alice Lee, whom I have most heartily to thank for her unwearying assistance, and once by myself. To preserve uniformity of working, the constants have in each case been carried to six figures. This involves little or no additional trouble, using as we do mechanical calculators. The final results are of course of no value beyond their probable errors, which will be in the second or third place of figures. No doubt I shall be told that there is a show of accuracy in the number of decimal figures retained, which does not really exist. It does not exist (and I am as fully conscious of its non-existance as any would-be critic) so far as our results fit the actual population, of which we have but a random sample. The figures, however, are of importance, as far as testing accuracy of fit of result to *actual* sample goes. The

* 'Biometrika,' vol. II., p. 12.

cubic or quartic curves may have coefficients insensible before the third or fourth figure of decimals, and these coefficients have to be multiplied occasionally by abscissæ of the third or fourth powers of 7 to 9. Hence to get ordinates true, *as far as the sample goes*, to the second or third figure, we require to work to a fairly high number of figures. There is no magic in six figures, four or five would probably satisfy another worker, but they are easily read off the calculator we use, and if the constants had been tabled only to four or five, no reader would have been able to agree exactly, if he wished to test any of our results, even to three figures, with the final ordinates.

DIAGRAM I. SKEW CORRELATION IN ASPERULA ODORATA.

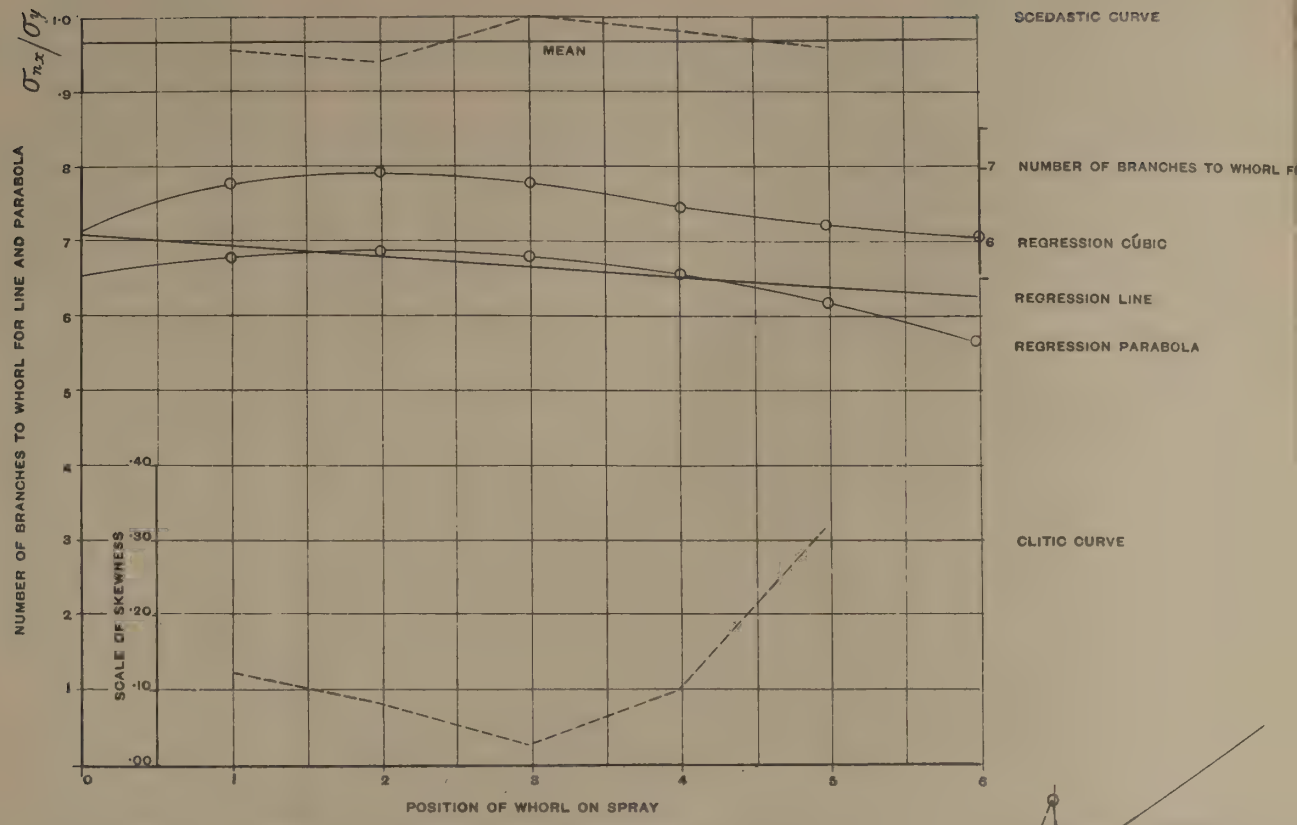


DIAGRAM II. SKEW CORRELATION, HEAD-HEIGHT AND AGE IN GIRLS.

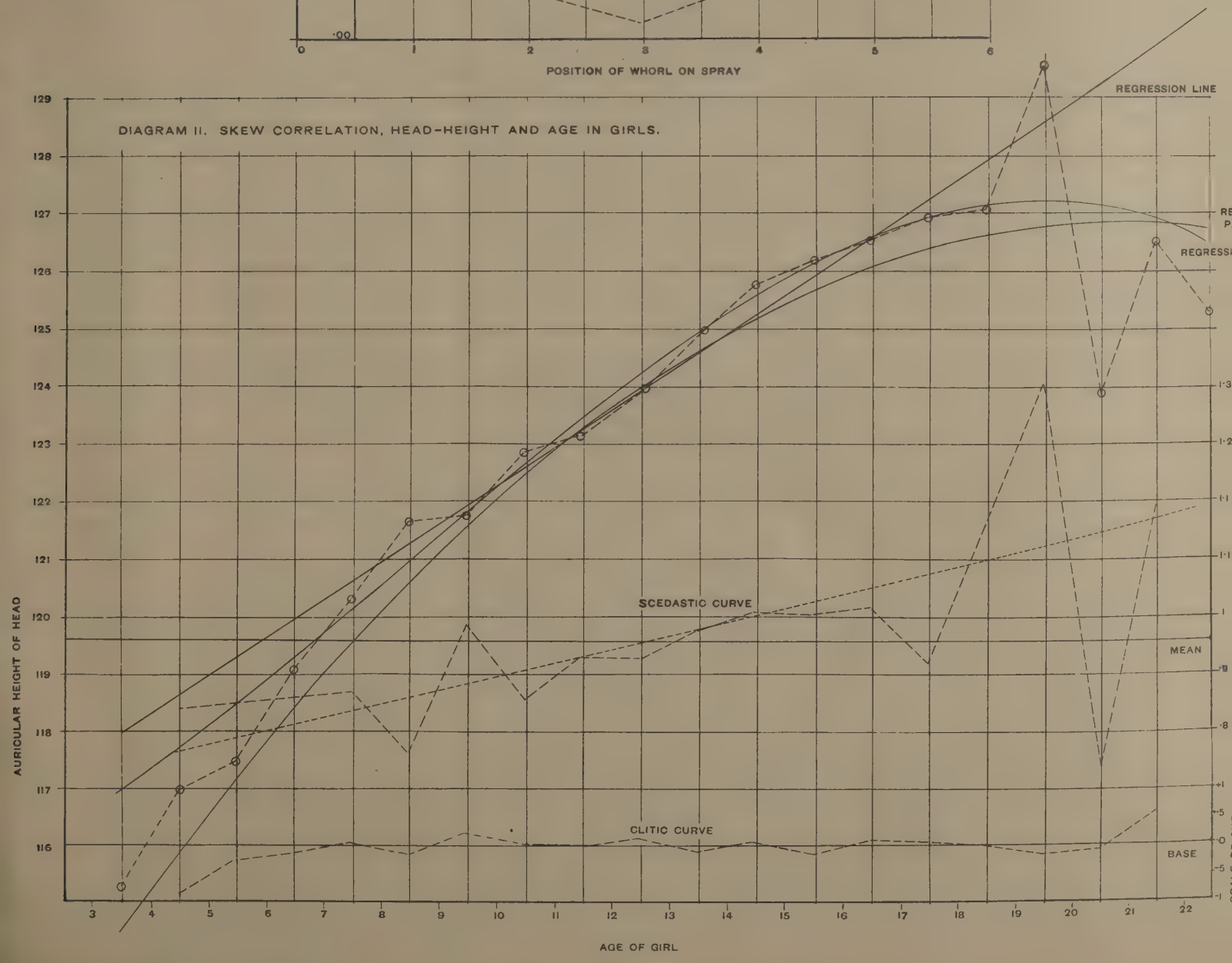


DIAGRAM III. SKEW CORRELATION BETWEEN SIZES OF CELL AND BODY IN DAPHNIA.

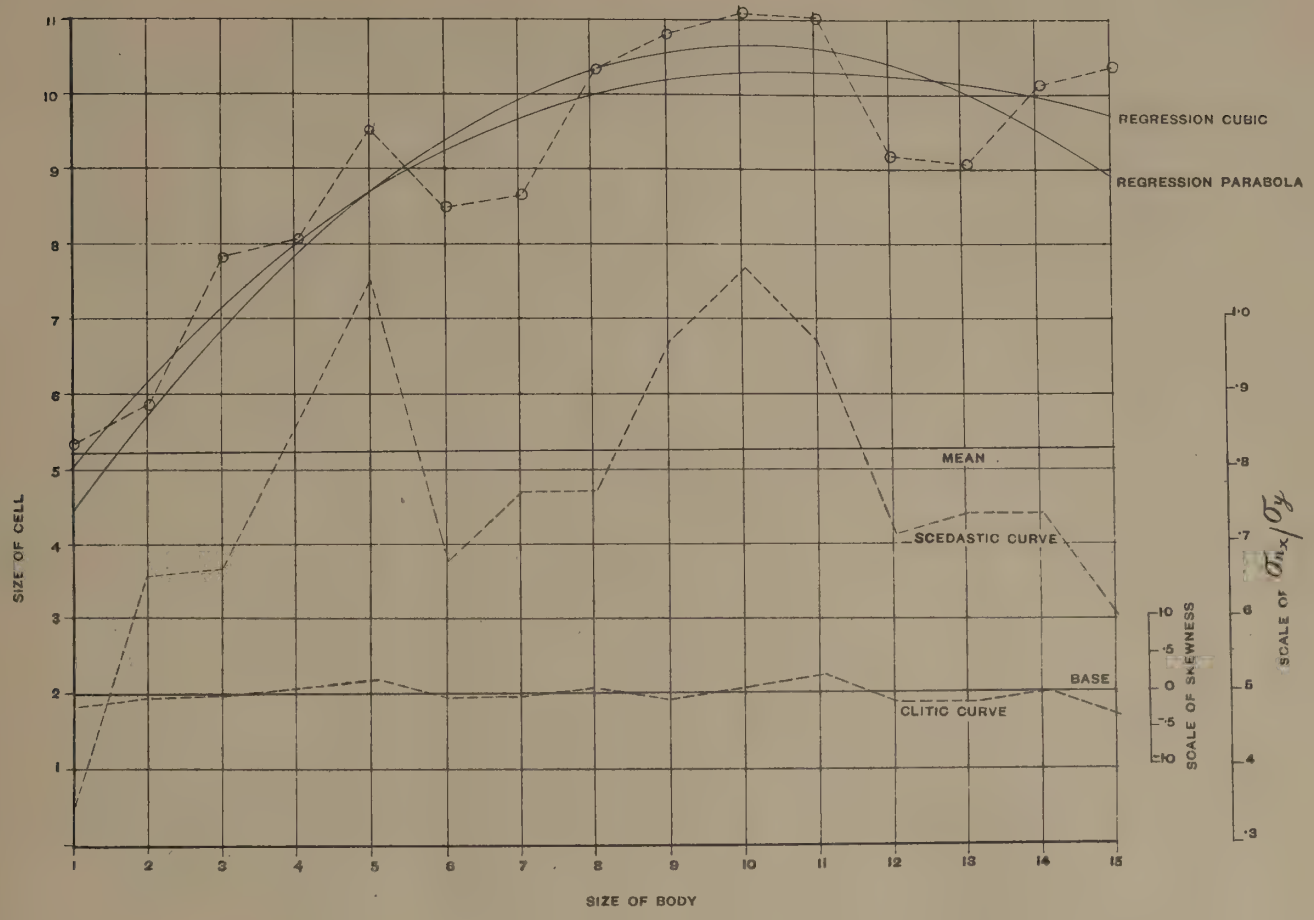


DIAGRAM IV. SKEW CORRELATION BETWEEN BRANCHES AND POSITION OF WHORL IN EQUISETUM: SCEDASTIC AND CLITIC CURVES

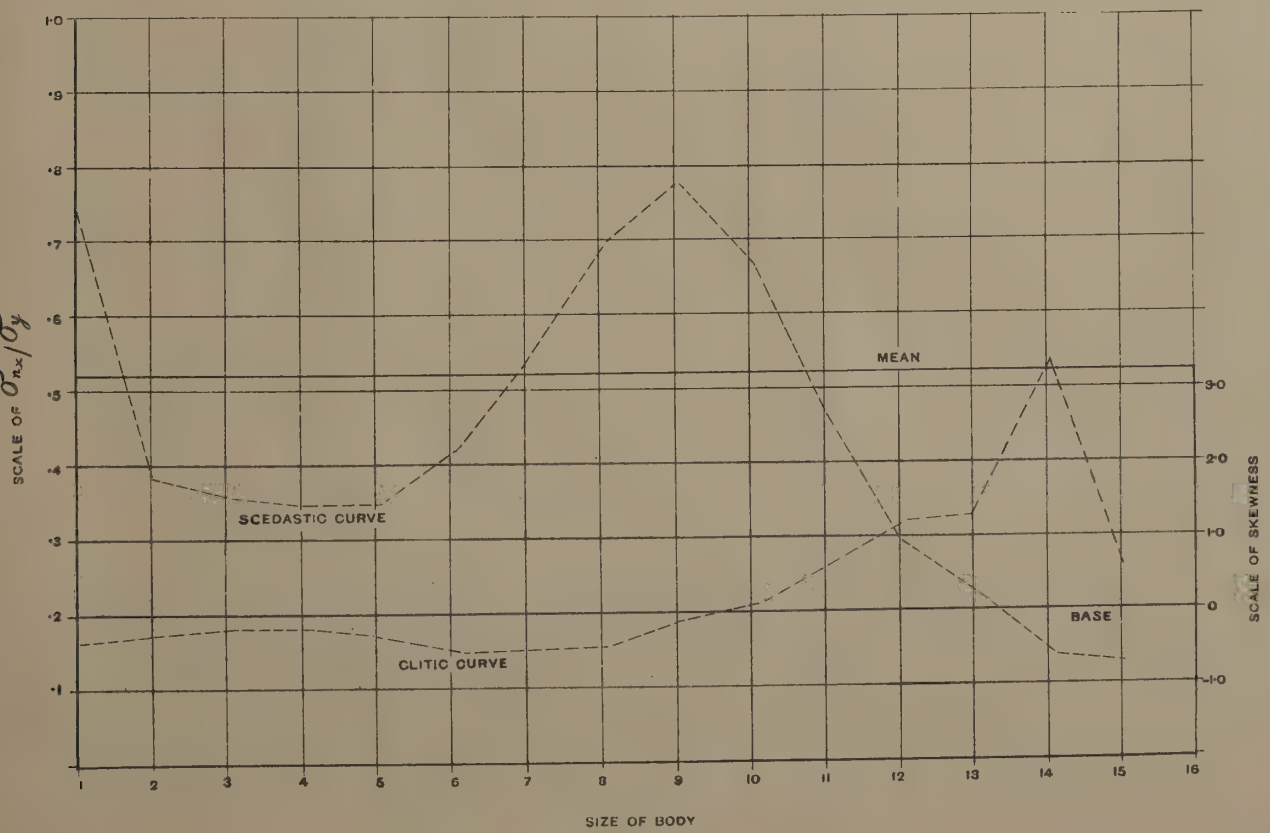
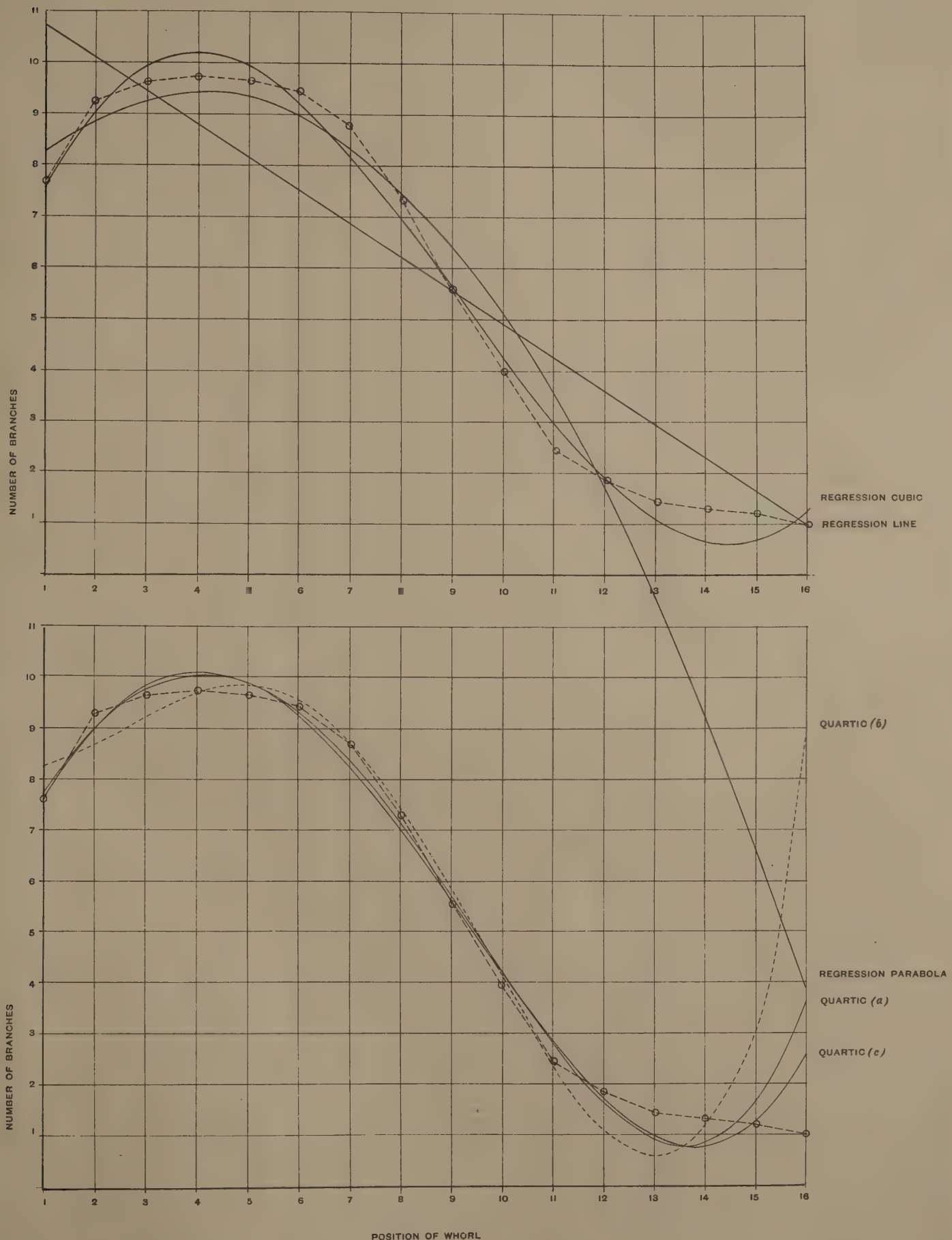


DIAGRAM V. SKEW CORRELATION BETWEEN BRANCHES AND POSITION OF WHORL IN EQUISETUM: REGRESSION CURVES.



DRAPERS' COMPANY RESEARCH MEMOIRS.

DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY COLLEGE, UNIVERSITY OF LONDON.

These memoirs will be issued at short intervals. The following are ready or will probably appear later in this series :—

Biometric Series.

I. Mathematical Contributions to the Theory of Evolution.—XIII. On the Theory of Contingency and its Relation to Association and Normal Correlation. By KARL PEARSON, F.R.S. *Issued.* Price 4*s.*

II. Mathematical Contributions to the Theory of Evolution.—XIV. On the Theory of Skew Correlation and Non-linear Regression. By KARL PEARSON, F.R.S. *Issued.* Price 5*s.*

III. Mathematical Contributions to the Theory of Evolution.—XV. On Homotyposis in the Animal Kingdom. By ERNEST WARREN, D.Sc., ALICE LEE, D.Sc., EDNA LEA-SMITH, MARION RADFORD and KARL PEARSON, F.R.S. *Shortly.*

Technical Series.

I. On a Theory of the Stresses in Crane and Coupling Hooks with Experimental Comparison with Existing Theory. By E. S. ANDREWS, B.Sc.Eng., assisted by KARL PEARSON, F.R.S. *Issued.* Price 3*s.*

II. On some Disregarded Points in the Stability of Masonry Dams. By L. W. ATCHERLEY, assisted by KARL PEARSON, F.R.S. *Issued.* Price 3*s.* 6*d.*

III. On the Graphics of Metal Arches, with Special Reference to the Relative Strength of Two-pivoted, Three-pivoted and Built-in Metal Arches. By L. W. ATCHERLEY and KARL PEARSON, F.R.S. *Issued.* Price 5*s.*

IV. On Torsional Vibrations in Shafting. By KARL PEARSON, F.R.S.

PUBLISHED BY DULAU AND CO.

MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.

XI. ON THE INFLUENCE OF SELECTION ON THE VARIABILITY AND CORRELATION OF ORGANS.

By KARL PEARSON, F.R.S.

'Phil. Trans.,' vol. 200, pp. 1–56. Price 3*s.*

XII. ON A GENERALISED THEORY OF ALTERNATIVE INHERITANCE, WITH SPECIAL REFERENCE TO MENDEL'S LAWS.

By KARL PEARSON, F.R.S.

'Phil. Trans.,' vol. 203, pp. 53–86. Price 1*s.* 6*d.*

PUBLISHED BY THE CAMBRIDGE UNIVERSITY PRESS.

BIOMETRIKA.

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS.

Edited, in Consultation with FRANCIS GALTON,

By W. F. R. WELDON, KARL PEARSON and C. B. DAVENPORT.

VOL. III., PARTS II. AND III.

I. Experimental and Statistical Studies upon Lepidoptera. I. Variation and Elimination in *Philosamea cynthia*. By HENRY EDWARD CRAMPTON.

II. On the Laws of Inheritance in Man.—II. On the Inheritance of the Mental and Moral Characters in Man, and its Comparison with the Inheritance of the Physical Characters. By KARL PEARSON.

III. A Study of the Variation and Correlation of the Human Skull with Special Reference to English Crania. By W. R. MACDONELL. (With 50 Plates.)

IV. On the Inheritance of Coat-colour in the Greyhound. By AMY BARRINGTON, ALICE LEE and K. PEARSON.

V. Note on a Race of *Clausilia itala* (Von Martens). By W. F. R. WELDON.

Miscellanea. On an Elementary Proof of SHEPPARD's Corrections for Raw Moments and on some Allied Points. (Editorial.)

VOL. III., PART IV.

I. Merism and Sex in *Spinax niger*. By R. C. PUNNETT.

II. Note on Inheritance of Meristic Characters in *Spinax niger*. By K. PEARSON.

III. On the Measurement of Internal Capacity from Cranial Circumferences. By M. A. LEWENZ and K. PEARSON. (With two Plates.)

IV. Étude Biométrique sur les Variations de la Fleur et sur l'Hétérostylie de *Pulmonaria officinalis L.* Par EDMUND GAIN.

Miscellanea. (I.) On the Correlation between Hair Colour and Eye Colour in Man. By K. PEARSON.

(II.) On the Correlation between Age and the Colour of Hair and Eyes in Man. By G. UCHIDA.

(III.) On the Contingency between Occupations in the Case of Father and Son. By EMILY PERRIN.

(IV.) On a Convenient Means of Drawing Curves to various Scales. By G. UDNY YULE.

(V.) Albinism in Sicily. By W. BATESON.

The subscription price, payable in advance, is 30*s. net* per volume (post free); single numbers 10*s. net.* Volumes I. to III. (1902–4) complete, 30*s. net* per volume. Bound in Buckram 34*s.* 6*d. net* per volume. Subscriptions may be sent to Messrs. C. J. Clay & Sons, Cambridge University Press Warehouse, Ave Maria Lane, London, either direct or through any bookseller.